Worst Case Analyses of Nearest Neighbor Heuristic for Finding the Minimum Weight k - cycle

Tanapat Chalarux and Piyashat Sripratak*

Department of Mathematics, Faculty of Science, Chiang Mai University, Chiang Mai, Thailand

Received: 7 January 2020, Revised: 6 February 2020, Accepted: 12 February 2020

Abstract

Given a weighted complete graph (K_n , w), where w is an edge weight function, the minimum weight k - cycle problem is to find a cycle of k vertices whose total weight is minimum among all k - cycles. Traveling salesman problem (TSP) is a special case of this problem when $k = n$. Nearest neighbor algorithm (NN) is a popular greedy heuristic for TSP that can be applied to this problem. To analyze the worst case of the NN for the minimum weight k - cycle problem, we prove that it is impossible for the NN to have an approximation ratio. An instance of the minimum weight k - cycle problem is given, in which the NN finds a k - cycle whose weight is worse than the average value of the weights of all k - cycles in that instance. Moreover, the domination number of the NN when $k = n$ and its upper bound for the case $k = n - 1$ is established.

Keywords: minimum weight k **-** cycle, worst case analysis, nearest neighbor heuristic
DOI 10.14456/cast.2020.7

1. Introduction

Traveling salesman problem is one of the most famous problem in mathematics and computer sciences [1]. Let G be a graph and cycle C be a subgraph of G whose vertex set $V(C)$ are the same as $V(G)$, the vertex set of G. C is called a *Hamiltonian cycle* or *tour*. Given a complete graph K_n with a weight function w from the edge set of K_n to the set of positive real numbers, the *symmetric traveling salesman problem* (STSP) seeks for a tour which has the minimum total weight among all tours in the graph. The *asymmetric traveling salesman problem* (ATSP) is defined similarly to the STSP by given a directed complete graph $\overleftrightarrow{K}_n$ instead of a complete graph K_n . The TSP is well-known to be NP-hard [2, 3], so it is difficult to find an optimal solution of the TSP with large number of vertices. There are no polynomial time algorithms to solve either the ATSP or the STSP, unless P = NP. Since the problem has been introduced, many popular heuristics for constructing a tour for the TSP such as greedy heuristics [4], nearest neighbor heuristics [4, 5] and local search heuristics [4, 6] have been proposed.

*Corresponding author: Tel.: +66 61 267 6777
E-mail: psripratak@gmail.com

There are many generalizations of the TSP studied throughout the years. Gutin and Karapetyan [7] have considered the generalized traveling salesman problem. This generalized version is to find a minimum weight cycle C in K_n whose vertex set is partitioned into M partite sets and C is composed of exactly one vertex from each partition set. Khachay and Neznakhina [8] have worked on the generalization of the TSP in the sense of the cycle cover problem (CCP).

In this research, we consider the *minimum weight k - cycle problem,* which is to find a minimum weight k - cycle among all k - cycles in a complete undirected graph K_n with weight function w for a fixed integer $k \le n$. When $k = n$, the minimum weight k - cycle problem and the traveling salesman problem are the same, so we can say that the minimum weight k - cycle problem is a generalization of the STSP. Hence, the minimum weight k - cycle problem is NP - hard.

Gutin *et al*. [9] have shown that greedy type heuristics are not appropriate for the TSP since for each greedy type heuristic they consider, there is an instance of the TSP such that that heuristic constructs a poor result. Precisely, they also show that the domination number of pure greedy heuristic and the NN are 1, and the domination number of the repetitive nearest neighbor heuristic for the STSP is at most 2^{n-3}.

However, when the size of a minimum weight cycle is not n, we cannot ensure that the domination number of the NN is still 1. In this work, we concentrate on the domination number of the NN for some specific values of k, and also consider some other aspects for worst case analyses, namely approximation ratio and no worse than average guarantee.

2. Materials and Methods

We evaluate the NN using three different methods: approximation ratio, no worse than average guarantee and domination number.

The approximation ratio is the most popular way to analyze a heuristic. Many studies use the approximation ratio to analyze heuristics for the TSP. Brecklinghaus and Hougardy [10] find the approximation ratio of the greedy algorithm for some special cases of TSP and that of the Clarke-Wright savings heuristic for the metric TSP. Nilsson [4] refers to the approximation ratio as an evaluation of some of tour construction heuristics and tour improvement heuristics.

We denote $G = (V(G), E(G))$ a graph with vertex set $V(G) = \{1, 2, 3, \dots, n\}$ and edge set $E(G)$. For any $u, v \in V(G)$, denote $e = \{u, v\}$ an edge from vertex u to vertex v. We call a pair of complete undirected graph K_n and its weight function w an *instance* (K_n, w) of the minimum weight k - cycle problem. Denote $w(u, v)$ a weight of an edge $\{u, v\}$ of graph G and $w(G)$ the sum of the weights of all edges in graph G. We denote $w(S)$ the maximum weight of a k - cycle constructed by heuristic and $w(T)$ the weight of a minimum k - cycle.

Definition 2.1 A heuristic A for the minimum weight k - cycle problem has the *approximation ratio* $\alpha \ge 1$ if for each instance with minimum k - cycle T, the heuristic A finds a k - cycle S such that $\dfrac{w(S)}{w(T)} \le \alpha$.

We also analyze the heuristic by checking whether it is worse than average. The idea of not worse than average heuristic is introduced in Russian literature. Punnen *et al*. [11] used average value based analysis on some heuristics for the bipartite boolean quadratic programming problem.

Definition 2.2 For each instance I, the average value of weights of all k - cycles is called the *average value* of I, and heuristic A is said to be *not worse than average* if heuristic A constructs a k - cycle of weight less than or equal to the average value of I for all instance I.

The domination number suggested by Glover and Punnen [12] is a new approach for evaluating heuristics. They propose a heuristic for the TSP with complexity $O(n)$ and give the domination number for that heuristic. Gutin *et al*. [9] studied the domination number of some greedy type heuristics for TSP.

For an instance of the minimum weight k - cycle problem, let H and S be k - cycles, we say that H *dominates* S if $w(H) \le w(S)$.

Definition 2.3 The *domination number* of a heuristic A for the minimum weight k - cycle problem on K_n is the maximum integer $d(n, k)$ such that, for any instance of the minimum weight k - cycle problem on n vertices, A produces a k - cycle K which dominates at least $d(n, k)$ cycles in I including K itself.

Among all greedy type heuristics studied in Gutin *et al*. [9], the *nearest neighbor heuristic* (NN) is a heuristic that we are interested since it can be directly applied to our problem. Let u be a vertex in graph G with vertex set $V(G)$. A vertex v is called *nearest vertex* from u if $w(u, v) = \min\{w(u, h): h \in V(G)\}$.

The NN starts constructing a cycle from a fixed vertex i_1, goes to i_2, which is the nearest vertex from i_1 ($w(i_1,i_2) = \min\{w(i_1, j): j \ne i_1\}$), and adds edge $\{i_1, i_2\}$, then to i_3, the nearest vertex from i_2 distinct from i_1 and i_2, and adds edge $\{i_2, i_3\}$. Repeat until we collect k vertices. Then add edge $\{i_k, i_1\}$.

In this work, we represent a path $P = (V(P), E(P))$ where $V(P) = \{v_1, v_2 \ldots, v_k\}$ and $E(P) = \{\{v_1, v_2\}, \{v_2, v_3\}, ..., \{v_{k-1}, v_k\}\}$ as $(v_1, v_2 \ldots, v_k)$. Denote P_n a path with $|V(P)| = n$. We say that P has length $n - 1$. The graph C constructed from a path P by adding an edge $\{v_k, v_1\}$ is called a *cycle*, denoted by $C = (v_1, v_2 \ldots, v_k, v_{k+1} = v_1)$.

3. Results and Discussion

Henceforth, denote $w(S)$ the maximum weight of a k - cycle constructed by the NN and $w(T)$ the weight of the minimum k - cycle.

In the following theorem, we show that approximation ratio is not appropriate for analyzing the NN for the minimum weight k - cycle problem on K_n . For any positive integer $n \ge 3$ and for each $\sigma \ge 1$, we can find an instance of the minimum weight k - cycle problem on K_n such that the ratio of $w(S)$ and $w(T)$ is greater than or equal to σ .

Theorem 3.1 Let n and k be positive integers where $n \ge 4$ and $3 \le k < n$. For any $\sigma \ge 1$, there exists an instance of the minimum weight k - cycle problem on complete graph K_n such that $\dfrac{w(S)}{w(T)} \ge \sigma$.

Proof Let $\sigma \geq 1$ and $3 \leq k < n$. The following instance of the minimum weight k - cycle problem on K_n is considered. Assume that edges $\{i, i + 1\}$ for $1 \leq i \leq n-1$ and edge $\{n, 1\}$ have weight σ, edge $\{k, 1\}$ has weight $\sigma^2(k+1)-(k-1)\sigma$, and all remaining edges have weight 2σ. Applying the NN starting at vertex 1, the NN constructs the k - cycle $C = (1, 2, 3, \ldots, k, 1)$. Then

$$w(S) \geq w(C) = \sigma(k-1) + (\sigma^2(k+1) - \sigma(k-1)) = \sigma^2(k+1).$$

Next, we show that k - cycle $T = (1, 2, 3, \ldots, k-1, n, 1)$ is a minimum k - cycle. Suppose that H is a k - cycle such that $w(H) < w(T)$. Note that edges of weight σ are the cheapest edges. The weight of $\{k, 1\}$ is $\sigma^2(k+1)-(k-1)\sigma > 2\sigma$. Moreover, T contains $k-1$ edges of weight σ and one edge of weight 2σ. Then H contains k edges of weight σ, which is impossible.

Since $T = (1, 2, 3, \ldots, k-1, n, 1)$ is the minimum k - cycle, we have $w(T) = \sigma(k-1) + 2\sigma = \sigma(k+1)$.

Therefore, $\dfrac{w(S)}{w(T)} \geq \dfrac{\sigma^2(k+1)}{\sigma(k+1)} = \sigma$.

In case of $k = n$, there exists an instance of the minimum weight k - cycle problem defined as follows:

Assume that any edge $\{u, u + 1\}$ where $1 \leq u \leq n-1$ and the edge $\{n, 1\}$ have weight σ, the edge $\{n-3, n-1\}$ has weight $\sigma - 1$, the edge $\{n-2, n\}$ has weight $\sigma(n\sigma - n + 2)$, and all remaining edges have weight 2σ. It is possible that the NN starting at 1 constructs k - cycle $C = (1, 2, \ldots, n-3, n-1, n-2, n, 1)$ with $w(C) = (n-1)\sigma - 1 + \sigma(n\sigma - n + 2) = \sigma^2 n + \sigma - 1$.

Next, we claim that the tour $T = (1, 2, \ldots, n, 1)$ is a minimum k - cycle with $w(T) = n\sigma$. Suppose that H is a k - cycle such that $w(H) < w(T)$. Thus, H is composed of the only edge that has weight less than σ, which is $\{n-3, n-1\}$ of weight $\sigma - 1$, and a Hamiltonian path P from $n-3$ to $n-1$.

If there is an edge of weight 2σ or $\sigma(n\sigma - n + 2) > 2\sigma$ in P, then

$$w(H) = w(n-3, n-1) + w(P) \geq \sigma - 1 + (n-2)\sigma + 2\sigma = n\sigma + \sigma - 1 > n\sigma = w(T),$$

a contradiction. Hence, all edges in P are of weight σ. Therefore, $n-3$ is adjacent to $n-4$ or $n-2$, and $n-1$ is adjacent to $n-2$ or n. Since $k = n \geq 4$, $n-2$ cannot be adjacent to both $n-3$ and $n-1$. Hence, an edge incident to $n-2$ has weight at least 2σ, a contradiction. Therefore, T is a minimum k - cycle.

Hence, $\dfrac{w(S)}{w(T)} \geq \dfrac{\sigma^2 n + \sigma - 1}{n\sigma} > \sigma.$

In the next proposition, we find a general formula of the average value of the weights of all k - cycles for any instance.

Theorem 3.2 Let n and k be positive integers where $n \geq 3$ and $3 \leq k \leq n$. For an instance of the minimum weight k - cycle problem (K_n, w), the average value of the weights of all k - cycles is $\dfrac{2k}{n(n-1)} \sum_{e \in E(K_n)} w(e)$.

Proof We find the number of all k - cycles in a complete graph K_n. There are $\binom{n}{k}$ ways to choose k vertices from n vertices to be in a k - cycle. Since the number of cycles composed of distinct k vertices is $\frac{(k-1)!}{2}$, the number of all k - cycles in a complete graph K_n is $\binom{n}{k}\frac{(k-1)!}{2}$. Next, we find the summation of weights of all k - cycles. We consider an arbitrary edge $\{a, b\}$. To construct a k - cycle, we find a path of k vertices from a to b. There are $\binom{n-2}{k-2}(k-2)!$ possible paths of k vertices from a to b. Thus, the summation of weights of all k - cycles is $\binom{n-2}{k-2}(k-2)!\sum_{e\in E(K_n)} w(e)$.

The average value of weights of all k - cycles is

$$\frac{\binom{n-2}{k-2}(k-2)!\sum_{e\in E(K_n)} w(e)}{\binom{n}{k}\frac{(k-1)!}{2}} = \frac{2k}{n(n-1)}\sum_{e\in E(K_n)} w(e).$$

Hence, the average value of the weights of all k - cycles is $\frac{2k}{n(n-1)}\sum_{e\in E(K_n)} w(e)$.

Next, we show that the NN for the minimum weight k - cycle problem can be worse than average by constructing an instance such that the NN constructs a cycle of weight greater than the average value of the weights of all k - cycles in that instance. In the case that $k = n = 3$, the NN constructs the optimal solution for all instances since there is only one possible k - cycle in each instance. Then we consider the case when $n \geq 4$ and $3 \leq k \leq n$.

Theorem 3.3 Let n and k be positive integers where $n \geq 4$ and $3 \leq k \leq n$. There is an instance such that the NN for the minimum weight k - cycle problem is worse than average.

Proof Consider an instance of the minimum weight k - cycle problem on K_n such that all edges have weight 1 except edge $\{1, k\}$ where $w(1,k) = 1+\frac{n}{2}(n-1)$. The average value of this instance is

$$\frac{2k}{n(n-1)}\{\frac{n}{2}(n-1)-1+\frac{n}{2}(n-1)+1\} = 2k.$$

Apply the NN starting at vertex 1. One of the cycles that we can obtain from the NN is the k - cycle $S = (1, 2, 3, \ldots, k, 1)$ with weight $k-1+1+\frac{n}{2}(n-1) = k+\frac{n}{2}(n-1) > 2k$. Thus, the NN for the minimum weight k - cycle problem is worse than average.

We consider the domination number of the NN for $k = n - 1$ and $k = n$. We first show an upper bound of the domination number of the NN for the case $k = n - 1$.

Theorem 3.4 Let n be a positive integer where $n \geq 4$. The domination number of the NN for the minimum $(n - 1)$ - cycle problem on K_n is at most $\frac{1}{2}(n-2)!+1$.

Proof For any edge $\{i, j\} \in E(K_n)$, we define weight function w as follow:

$$w(i, j) = \begin{cases} in; & 1 \le i \le n-2, \ j = i+1 \\ in+1; & 2 \le i \le n-2, \ j \ge i+2 \\ n+1; & i = 1, 3 \le j \le n-2 \text{ or } j = n \\ (n-1)(n-2)n+1; & i = 1, \ j = n-1 \\ 1; & i = n-1, \ j = n. \end{cases}$$

Suppose that the NN starts at vertex 1. Then the $(n – 1)$ – cycle constructed by the NN is $C_{NN} = (1, 2, 3, \ldots, n-1, 1)$ where $w(C_{NN}) = n\sum_{i=1}^{n-2} i + (n-1)(n-2)N + 1$. We consider an $(n – 1)$ - cycle H that does not contain edge $\{1, n – 1\}$. Since

$$w(1, n-1) > (n-1) \cdot max\{w(i, j) \mid \{i, j\} \in E(K_n) \setminus \{\{1, n-1\}\}\}, w(H) < w(C_{NN}).$$

We can see that the number of $(n – 1)$ - cycles which do not contain edge $\{1, n – 1\}$ is

$$\frac{n}{2}(n-2)! - (n-2)!.$$

In the next step, we consider $(n – 1)$ - cycle $C \ne C_{NN}$ which contains edge $\{1, n – 1\}$. This $(n – 1)$ - cycle C is composed of an edge $\{1, n – 1\}$ and the path P of length $n – 2$ starting from 1 and ending at $n – 1$, where $P = (1 = v_1, v_2, v_3, \ldots, v_{n-1} = n-1)$. Assume that C does not contain vertex n. Let $B = \{\{v_i, v_{i+1}\} \in E(P) \mid v_i > v_{i+1}\}$. Since $C \ne C_{NN}$, B is not empty. Then

$$w(C) \le w(C_{NN}) + N \sum_{\{v_i, v_{i+1}\} \in B} (v_{i+1} - v_i) + \varepsilon(n-2) < w(C_{NN}).$$

Thus, the $(n – 1)$ - cycle $C \ne C_{NN}$, which contains edge $\{1, n – 1\}$ and does not contain vertex n, is not dominated by C_{NN}. The number of these $(n – 1)$ - cycles is $(n – 3)! – 1$.

Next, we consider the case when C contains vertex n. Assume that $v_i = n$. Let u be a vertex that does not show in $(n – 1)$ - cycle C and $B' = \{\{v_j, v_{j+1}\} \in E(P) \mid v_j > v_{j+1}, v_j \ne n\}$. Then

$$w(C) \le w(C_{NN}) + (v_{i+1} - u)N + \sum_{\{v_j, v_{j+1}\} \in B'} (v_{j+1} - v_j)N + (n-2)\varepsilon.$$

Since $\sum_{\{v_j, v_{j+1}\} \in B'} (v_{j+1} - v_j)N \le 0$, we have $w(C) \le w(C_{NN}) + (v_{i+1} - u)N + (n-2)\varepsilon$. If $v_{i+1} - u < 0$, we can conclude that $w(C) < w(C_{NN})$. Hence, we count the number of $(n – 1)$ - cycles which contain edge $\{1, n – 1\}$ and vertex n where $v_{i+1} < u$. We see that the number of ways to choose vertices u and v_{i+1} satisfying the condition $u > v_{i+1}$ is $\frac{1}{2}(n-3)(n-4)$. Since the path P is a sequence of n – 1 vertices starting with $v_1 = 1$ and ending with $v_{n-1} = n-1$, we just need to fill in the remaining n – 3 positions by the $n – 3$ vertices other than vertices 1, u and $n – 1$, so that n is followed by the given v_{i+1}. There are $(n – 4)!$ ways to complete this step. Thus, the number of ways to construct $(n – 1)$ - cycle satisfying the condition is $\frac{1}{2}(n-3)!(n-4)$.

From all cases, there are at least

$$\left(\frac{n}{2}(n-2)! - (n-2)!\right) + \left((n-3)! - 1\right) + \left(\frac{1}{2}(n-3)!(n-4)\right) = \frac{n}{2}(n-2)! - \frac{n-2}{2}(n-3)! - 1$$

$(n-1)$ - cycles that are not dominated by C_{NN}. Note that the number of all $(n-1)$ - cycles in K_n is $\frac{n}{2}(n-2)!$. Then C_{NN} can dominate at most $\frac{n}{2}(n-2)! - (\frac{n}{2}(n-2)! - \frac{n-2}{2}(n-3)! - 1) = \frac{1}{2}(n-2)! + 1$ $(n-1)$ - cycles including itself. Hence, the domination number of the NN for the minimum weight $(n-1)$ - cycle problem is at most $\frac{1}{2}(n-2)! + 1$.

We can slightly modify the instance used in the proof of Theorem 3.4 to give an instance in which the NN gives the unique maximum weight *k* - cycle. As a result, the domination number of the NN for the minimum weight *k* - cycle problem for *k* = *n* is 1. Since our problem becomes the STSP when *k* = *n*, the next theorem offers a verification for the domination number of the NN for the STSP, which is mentioned without proof by Gutin *et al*. [9]. Hence, the proof is omitted.

Theorem 3.5 Let *n* be a positive integer where $n \geq 3$. The domination number of the NN for the minimum weight *n* - cycle problem on K_n is 1.

4. Conclusions

We point out that approximation ratio is not an appropriate method to analyze the NN for the minimum weight *k* - cycle problem. We show that for any $\sigma \geq 1$, there is an instance of the minimum weight *k* - cycle problem such that the ratio of the maximum weight of *k* - cycle constructed by the NN and the weight of a minimum *k* - cycle is greater than or equal to σ.

Secondly, we find the average value of the weights of all *k* - cycles for each instance, and construct an instance of the minimum weight *k* - cycle problem such that the weight of a *k* - cycle constructed by the NN is greater than the average value of the weights of all *k* - cycles in the instance. Thus, the NN for the minimum weight *k* - cycle problem can be worse than average.

Finally, we establish an upper bound $\frac{1}{2}(n-2)! + 1$ for the domination number of the NN for the minimum $(n-1)$ - cycle problem. Moreover, we prove that the domination number of the NN for the STSP is 1.

Even the NN can give the unique worst solution for some instance of the TSP, the domination analysis shows that it is more promising when $k \neq n$. The output from the NN can be used as an initial solution in more complicated heuristics to get a better solution.

References

[1] Lawler, E.L., Lenstra, J.K., Rinnooy Kan, A.H.G. and Shmoys, D.B., 1985. *The Traveling Salesman Problem*. Chichester: John Wiley and Sons Ltd.

[2] Gutin, G., 2001. TSP tour domination and Hamilton cycle decomposition of regular digraphs. *Operations Research Letters*, 28, 107-111.

[3] Gutin, G., Punnen, A.P. and Lenstra, J.K., 2002. *The Traveling Salesman Problem and Its Variations*. Boston: Kluwer Academic Publishers.

[4] Nilsson, C., 2003. *Heuristics for the traveling salesman problem*. [online] Available at: http://160592857366.free.fr/joe/ebooks/ShareData/Heuristics%20for%20the%20Traveling%20Salesman%20Problem%20By%20Christian%20Nillson.pdf

[5] Khan, A.A. and Agrawala, H., 2016. Comparative study of nearest neighbour algorithm and genetic algorithm in solving traveling salesman problem. *International Research Journal of Engineering and Technology*, 3, 234-238.

[6] Wu, Y., Weise, T. and Chiong, R., 2015. Local search for the traveling salesman problem: a comparative study. *Proceedings of IEEE 14th International Conference on Cognitive Informatics and Cognitive Computing*, 213-220.

[7] Gutin, G. and Karapetyan, D., 2009. Generalized traveling salesman problem reduction algorithms. *Algorithmic Operations Research*, 4, 144-154.

[8] Khachay, M. and Neznakhina, K., 2016. Approximability of the minimum-weight k -size cycle cover problem. *Journal of Global Optimization*, 66, 65-82.

[9] Gutin, G., Yeob, A. and Zverovich, A., 2002. Traveling salesman should not be greedy: domination analysis of greedy-type heuristics for the TSP. *Discrete Applied Mathematics*, 117, 81-86.

[10] Brecklinghaus, J. and Hougardy, S., 2015. The approximation ratio of the greedy algorithm for the metric traveling salesman problem. *Operations Research Letters*, 43, 259-261.

[11] Punnen, A.P., Sripratak, P. and Karapetyan, D., 2015. Average value of solutions for the bipartite boolean quadratic programs and rounding algorithms. *Theoretical Computer Science*, 565, 77-89.

[12] Glover, F. and Punnen, A.P., 1997. The traveling salesman problem: new solvable cases and linkages with the development of approximation algorithms. *Journal of the Operational Research Society*, 48, 502-510.