Average TSP tour length approximations for territory design

Daisuke Hasegawa[1[0000-0002-4854-6665]] and Naoshi Shiono[2[0000-0003-0913-821X]]

[1] The University of Tokyo, Bunkyo, Tokyo, Japan
hasega60@e.u-tokyo.ac.jp
[2] Kanagawa Institute of Technology, Atsugi, Kanagawa, Japan
na-shiono@ic.kanagawa-it.ac.jp

Abstract. In vehicle routing, estimating route lengths using continuous approximation models can be valuable for delivery planning, especially for tour cost estimation and territory design because it avoids the computational cost associated with solving TSP and VRP directly. In this study, we propose a route length estimation formula based on rectilinear distances by considering the shape of the area. We calibrated the parameters through numerical experiments. Thus, our proposed formula can estimate the average tour length in rectilinear distance with high accuracy; and as the number of points increases, the influence of the shape decreases.

Keywords: Traveling salesman problem, Continuous approximation models, Rectilinear metric.

1 Introduction

In logistics, Japanese suppliers often designate separate of pickup and delivery territories for each driver. A reason to set territories is that customers request the same drivers to delivery their shipments. Another reason is that drivers travel more efficiently and safely because they are familiar with the characteristics of their routes and customers.

Therefore, the manager periodically reviews their territories to minimize the total working time for all drivers. For example, in Japan, an average customer replaces their gas cylinder 10 times a year. Then, for a driver, the customers to deliver cylinders to in each trip are different, i.e., the customers are stochastic. A typical driver changes 30–40 cylinders per trip.

To design the territories, we estimated the average tour length in the rectangle for a small number of customers (less than 100 customers) by continuous approximation approach and the rectilinear metric, where the origin is located at the corner of the rectangle.

Finding the shortest travel distance is well-known as the traveling salesman problem (TSP) [1] and vehicle routing problem (VRP). With the improvement in computer performance and algorithms to solve the problems, the size of problems that can be optimally solved have increased. However, it still requires multiple calculations in cases when the number of demand points is too large, when the exact locations of the demand points are unknown, or when the locations of demand points change every day. In such cases, a continuous approximation model can estimate the mean distance with high accuracy. Many previous studies have approximated tour distances using the Euclidian distance metric [2]. However, we focus on using the rectilinear distance metric that considers horizontal and vertical differences as distances.

The reason we apply the rectilinear metric is because it is easy to use in practice. Fig. 1 is an example of a territory design. If we assume the Euclidean metric among customers and depot, we need to estimate the line-haul distance, which is the distance from the depot to the nearest customer in the rectangle. However, the accuracy of the tour length estimation will sometimes be worse. Therefore, in this study, we use rectilinear distance estimation. This distance metric imitates movement on a real city street network (e.g., NYC, Kyoto, Barcelona) and does not require distance estimation for line-haul distances. After calculating the distance from the depot to the origin, the nearest corner of the rectangle, we add the calculated distance and the estimated tour length for visiting all customers and returning to the origin in the rectangle.

The rest of this paper is organized as follows. In Section 2, a literature review of TSP length using the continuous approximation approach is provided. We have explained the estimation procedure in Section 3. In Section 4, we describe the experimental design and validation results. Finally, in Section 5, we conclude the paper.

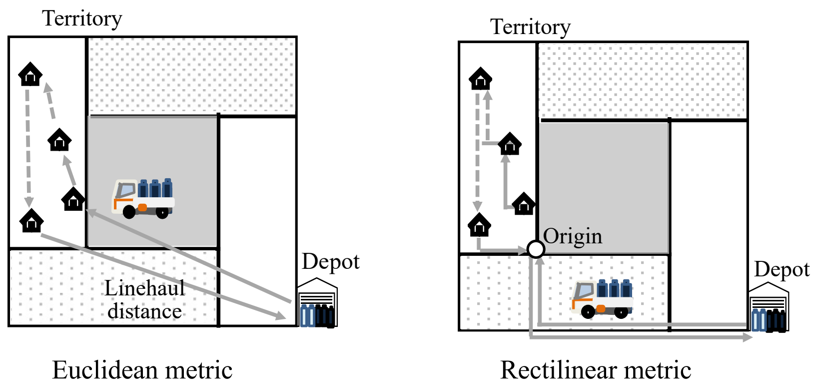

Fig. 1. Tour from depots to stochastic customers.

2 Literature review

We assume that n points are uniformly and independently distributed on the planar area of size A throughout the study. Then, models to find the optimal length are generally referred to the continuous approximation models. The need to approximate the optimal length has been studied since the 1940s. Franceschetti et al. [3] reviewed the approximation models in freight transportation management. They also suggested that most applications of continuous approximation models are categorized by districting, location, fleet composition, and vehicle routing. Districting and territory design for our objective are used with the same meanings.

There are many previous studies that assume the area to be the unit square $[0,1] \times [0,1]$. Let $T_{L2}(n)$ denote the approximate TSP tour length, the approximate length of the shortest tour covering all points, in the Euclidean distance. After Bearwood et al. [4] showed that

$$\lim_{n \to \infty} \frac{T_{L2}(n)}{\sqrt{n}} = \beta_{L2}, \tag{1}$$

with probability one for a constant β_{L2}, many researchers gave different estimates of β_{L2} using various methodologies [5]. Some researchers also estimated β_{L1}, the constant in the rectilinear metric. Johnson et al. [6] estimated β_{L1} to be 0.8943 ± 0.0007 while β_{L2} was 0.7124 ± 0.0002. However, they do not show the details to compute β_{L1}. Vig and Paleker [5] showed that

$$T_{L1}(n) = 0.765 + 0.892\sqrt{n}, \tag{2}$$

and

$$T_{L2}(n) = 0.662 + 0.710\sqrt{n} \tag{3}$$

using ordinary least squares (OLS) regression for 11 to 2000 points and proposed $\beta_{L1} = 0.892$ and $\beta_{L2} = 0.710$.

Some studies also focused on the difference in distance between two points, e.g., Euclidean, rectilinear, and Karlsruhe distances [7]. Consider U_1, U_2 and R be the rectilinear, Euclidean, and Karlsruhe distance between two points, respectively. Kobayashi and Koshizuka [8] showed that the ratio of the expected value of the Euclidean distance $E(U_2)$ to the value of the rectilinear distance $E(U_1)$ in the rectangle region is approximated by

$$E(U_1)/E(U_2) \sim 1.280, \tag{4}$$

if the length–width ratio of the region is 2 or less. Note that length and width are defined such that the ratio is always greater than or equal to 1. Additionally, Kurita [9] showed that the ratio in the circle region is expressed as:

$$E(U_1)/E(U_2) \cong 1.273 \tag{5}$$

and

$$E(R)/E(U_2) \cong 1.097. \tag{6}$$

Çavdar and Sokol [10] proposed the tour length formula for the dispersion of any points in the Euclidean metric. Points are scattered in a rectangle with central horizontal and vertical axes. Using regression by the tour length for 3000–8000 points in the 27 rectangles, they proposed the tour length formula $T'_{L2}(n)$ as:

$$T'_{L2}(n) = 2.791\sqrt{n(\sigma_{lx}\sigma_{ly})} + 0.267\sqrt{\frac{nA(\sigma_x\sigma_y)}{\bar{c}_x\bar{c}_y}}, \tag{7}$$

where $\bar{c}_x$ and $\bar{c}_y$ are average distance of points to the central horizontal and vertical axes, σ_x and σ_y are the standard deviation of the horizontal and vertical coordinates of the points, σ_{lx} and σ_{ly} are the standard deviation of the absolute distances of the points from the central horizontal and vertical axes, and A is the area of a rectangle.

Meanwhile, some researchers suggest that n is small in actual delivery. Vig and Paleker [5] showed that repair persons travel between 15 and 20 stops on a single trip. Holguín-Veras and Patil [11] suggested that 35% of all trip chains have only two stops and the maximum number of stops in a trip chain is 23 for commercial vehicles in Denver. In Japan, the person for periodic inspection for gas appliances travels at most 10 stops. The home delivery driver for an online supermarket travels at most 20 stops. Choi and Schonfeld [12] present a comprehensive review for approximating the TSP tour length for low n.

With regard to the approximation of the TSP tour length for small n in the rectangle, there are two well-known previous studies. Chein [13] estimated the TSP tour length using the Euclidean metric in eight rectangles whose length–width ratio λ is 1 to 8 and has eight sectorial-shaped regions. They produced 4160 instances in which 5–30 customers were randomly distributed, and the origin was one of the corners of the region. After seven tour length estimators are considered and evaluated, the tour length $t^*_{L2}(\lambda, n)$ is proposed as:

$$t^*_{L2}(\lambda, n) = 2.10\bar{r} + 0.67\sqrt{(n-1)R}, \tag{8}$$

where $\bar{r}$ is the average Euclidean distance from the customers to the depot and R is the area of the smallest rectangle that covers only the customers. $t^*_{L2}(\lambda, n)$ was obtained with high accuracy (Adjusted $R^2 = 0.99$). Additionally, they presented

$$t^*_{L2}(\lambda, n) = a_1\bar{r} + a_2\sqrt{(n-1)R}, \tag{9}$$

in the rectangle corresponding to λ, as shown in Table 1.

Table 1. Results of $t^*_{L2}(\lambda, n)$.

Coefficient	$t^*_{L2}(1,n)$	$t^*_{L2}(2,n)$	$t^*_{L2}(3,n)$	$t^*_{L2}(4,n)$
a_1	2.39	2.26	2.25	2.41
a_2	0.57	0.59	0.58	0.54

Coefficient	$t^*_{L2}(5,n)$	$t^*_{L2}(6,n)$	$t^*_{L2}(7,n)$	$t^*_{L2}(8,n)$
a_1	2.45	2.38	2.53	2.57
a_2	0.54	0.57	0.54	0.54

Kwon et al. [2] also estimated the tour length in the rectangle whose length–width ratio λ is 1 to 8 and whose customers are randomly distributed from 10 to 89. Then, the TSP tour length $t^*_{L2}(\lambda, n)$ of the rectangle was proposed as

$$t^*_{L2}(\lambda, n) = \left(0.833 - 0.001n + \frac{1.115\lambda}{n}\right)\sqrt{nA}, \tag{10}$$

or

$$t^*_{L2}(\lambda, n) = 0.414\bar{r} + \left(0.775 - 0.0008n + \frac{0.903\lambda}{n}\right)\sqrt{nA}, \tag{11}$$

where A is the area of the rectangle with high accuracy ($R^2 = 0.998$). In this study, our analysis is similar to that of Kwon et al. [2], but we assume the rectilinear metric during the tour.

3 Our approach

3.1 Experimental Procedures

There are many procedures to approximate the TSP tour length. In particular, there are two main methods. One method is to perform OLS regression on all the data, whereas the other method is to summarize the data, e.g., by averaging and then performing OLS regression. Considering of our objective, we should seek the average TSP tour length in the rectangle. Our procedure is based on that of Choi and Schonfeld [12].

Let N be the set of number of demand points (customers) and the origin. First, $|N| - 1$ demand points are generated according to a given distribution in a rectangle whose area is one. Let Λ be the set of length–width ratio. For $\lambda \in \Lambda$, by generating two random numbers uniformly distributed in the interval $[0, \sqrt{\lambda}]$ and $[0, 1/\sqrt{\lambda}]$, respectively, the numbers are regarded as x- and y- coordinates of a demand point in the rectangle. Each demand point in the rectangle $[0, \sqrt{\lambda}] \times [0, 1/\sqrt{\lambda}]$ is equally likely to be selected. Additionally, one point is generated in the origin, the corner of the rectangle.

Second, we select a solution method to compute the rectilinear TSP tour length for $n \in N$ points. In this study, we compute the rectilinear tour length using Helsgaun's implementation of the Lin–Kernighan heuristic algorithm [14]. Vig and Palekar [5] also applied this algorithm because Johnon and McGeoch [15] showed that the solutions from this algorithm were typically very close to the optimal solution.

Third, instances on a given n points were produced. Let I be the set of instances. After $|I|$ iterations for each n, the rectilinear tour lengths for each n points were averaged. Thereafter, the repeated runs move other $n' \in N$. Finally, the averaged rectilinear tour length was fitted with OLS regression.

3.2 Features of the tour length

In this study, we assume that

- $N = \{11, 21, \ldots, 91, 101\}$,
- $\Lambda = \{1, 2, 3, 4, 5\}$,
- $I = \{1, 2, \ldots, 499, 500\}$,

i.e., the maximum length–width ratio (area aspect ratio) is five and we have 500 iterations for each $n \in N$ and $\lambda \in \Lambda$.

Let $t_{\lambda,n,i}$ be the rectilinear tour length for $i \in I$, $n \in N$, and $\lambda \in \Lambda$. Then, the average tour length $\bar{t}_{\lambda,n}$ for $n \in N$ and $\lambda \in \Lambda$ is computed as:

$$\bar{t}_{\lambda,n} = \frac{1}{|I|} \sum_{i \in I} t_{\lambda,n,i}. \tag{12}$$

Fig. 2 shows the result of $\bar{t}_{\lambda,n}$. The result indicates that the average length is proportional to a concave function of n. Previous studies used the factor $\sqrt{n}$ and we consider applying $\sqrt{n}$ in the TSP tour approximation. Additionally, as the number of demand point increases, the effect of length–width ratio on the average length decreases. In fact, $\bar{t}_{5,101} - \bar{t}_{1,101} = 0.3$ while $\bar{t}_{5,11} - \bar{t}_{1,11} = 1.1$. Then, we have to consider the factor proportional to the length–width ratio and inversely proportional to the demand points.

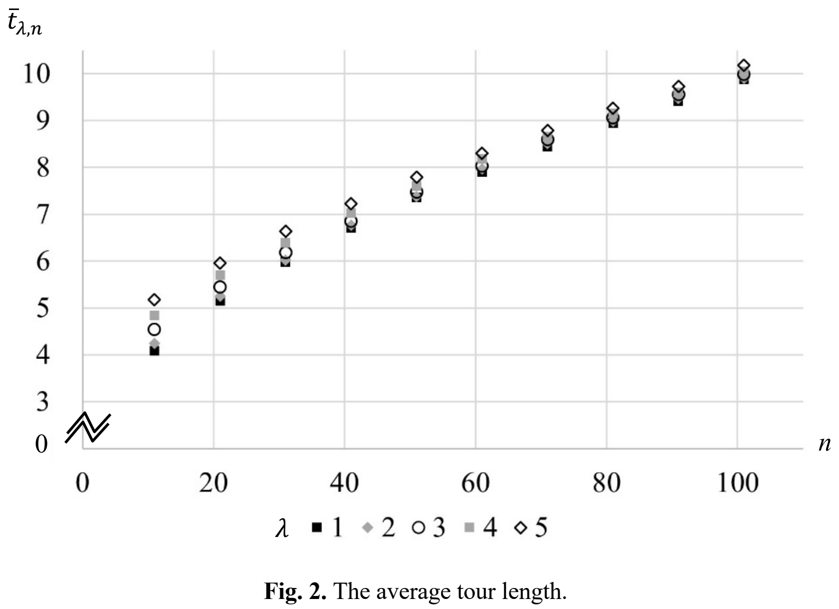

Fig. 2. The average tour length.

4 Regression based approximations

4.1 Formula in the unit rectangle

In this study, we investigated a simple approximation formula with high accuracy. As our objective is territory design, the tour length formula should be simple. Then, after running the OLS regression by the parameters used in Chien [13] and Kwong et al. [2] and our parameters, we have discovered the following TSP tour length approximation formula (Model. 1) as:

$$t_{L1}^{*}(\lambda, n) = a_1\sqrt{n} + \frac{a_2\lambda}{\sqrt{n}} + b, \tag{13}$$

where the area of the rectangle is one, the estimated parameters a_1, a_2 are 0.880, 0.872, and constraint b is 0.875, with Adjusted $R^2 = 0.999$. Compared to Eqs. (10) and (13), Eq. (10) contains a term of $n\sqrt{n}$, while Eq. (13) contains a constant term. In this study, we apply three measures for the quality of estimation in addition to R^2. The mean percentage error (MPE) is defined as:

$$\text{MPE} = \frac{1}{|N||\Lambda|}\sum_{\lambda\in\Lambda, n\in N}\frac{\bar{t}_{\lambda,n} - t_{L1}^{*}(\lambda, n)}{\bar{t}_{\lambda,n}} \times 100\%, \tag{14}$$

the mean absolute percentage error (MAPE) as:

$$\text{MAPE} = \frac{1}{|N||\Lambda|}\sum_{\lambda\in\Lambda, n\in N}\frac{\left|\bar{t}_{\lambda,n} - t_{L1}^{*}(\lambda, n)\right|}{\bar{t}_{\lambda,n}} \times 100\%, \tag{15}$$

and the maximum absolute percentage error (MaxAPE) as:

$$\text{MaxAPE} = \max_{\lambda\in\Lambda, n\in N}\frac{\left|\bar{t}_{\lambda,n} - t_{L1}^{*}(\lambda, n)\right|}{\bar{t}_{\lambda,n}} \times 100\%. \tag{16}$$

Additionally, we consider another formula (Model. 2), the simplified form of Model. 1, as:

$$t_{L1}^{*}(\lambda, n) = a_1\left(\sqrt{n} + \frac{\lambda}{\sqrt{n}} + 1\right). \tag{17}$$

The statistics for Model. 1 are shown in Table 2 and three measures for Models. 1 and 2 are shown in Table 3. Table 2 shows the standard errors are small for all coefficients. In addition, the MPE, MAPE, and MaxAPE are small for both equations. Moreover, Model. 1 are with low measures. Particularly, MaxAPE is the important factor considering the territory design, the values for both formulas are 2.0% or less.

Table 2. Statistics of Model. 1

Parameter	Coefficient	Standard error	t value	Pr (>$\|t\|$)
a_1	0.880	0.0353	24.8	<0.001
a_2	0.872	0.0036	240.0	<0.001
b	0.875	0.0259	33.7	<0.001

Table 3. Fit comparison for two equations

Formula	MPE	MAPE	MaxAPE
Model. 1	0.006%	0.5%	1.8%
Model. 2	0.1%	0.5%	2.0%

Furthermore, for evaluation, we generate other $\bar{t}_{\lambda,n}$ from 200 instances each and compute three measures, as shown in Table 4. MPE for both formulas become worse, but the value is small. Therefore, these results indicate that our formulas perform appropriately.

Table 4. Measures for other instances.

Formula	MPE	MAPE	MaxAPE
Model. 1	0.05%	0.6%	1.9%
Model. 2	0.7%	0.6%	1.8%

4.2 Formula in the rectangle and applications

Finally, our experiment assumes that the area of the rectangle is one in this study. If the area of the rectangle is A, our approximation models become:

$$t_{L1}^{*}(\lambda, n) = \left(a_1\sqrt{n} + \frac{a_2\lambda}{\sqrt{n}} + b\right)\sqrt{A}, \tag{18}$$

$$t_{L1}^{*}(\lambda, n) = a_1\left(\sqrt{n} + \frac{\lambda}{\sqrt{n}} + 1\right)\sqrt{A}, \tag{19}$$

respectively.

As mentioned in the first section, our objective is to model an equation for territory design. For example, in Fig. 1, we consider a territory divided into m rectangles of equal size. As n and A become constant, we only calculate the distance from the depot to the origin for each rectangle and $0.872\lambda/\sqrt{n}$ in Eq. (18). Therefore, our approximation equation is easy to use in territory design. However, our equation assumes that λ ranges from 1–5.

5 Conclusion

In this study, we proposed a continuous approximation model for touring distance based on rectilinear distance metric, which has the advantage of reduced computational cost. To evaluate the accuracy of our model, we varied the number of tour points and aspect ratio of the region, and then conducted numerical experiments by simulation. Our results indicate that our model can accurately estimate tour distances. Additionally, we observed that the influence of the aspect ratio decreases as the number of tour points increase. Our results are not only relevant to cities with gridded streets, but also prove to be effective for developing delivery plans in various urban environments. In future works, we will extend our model to address multi-vehicle delivery (VRP) and explore area partitioning methods for practical delivery planning applications.

Acknowledgements
This work was supported by JSPS KAKENHI [Grant Numbers JP21K14314 and JP23K04269]. We appreciate their support.

References

1. Applegate, D.L., Bixby, R.E., Chvatál, V., Cook, W.J.: The Traveling Salesman Problem; A Computational Study. Princeton University Press, Princeton, New Jersey (2006).
2. Kwon, O., Golden, B., Wasil, E.: Estimating the length of the optimal TSP tour: An empirical study using regression and neural networks. Computers & Operations Research **22**(10), 1039–1046 (1995). doi: /10.1016/0305-0548(94)00093-N.
3. Franceschetti, A., Jabali, O., Laporte, G.: Continuous approximation models in freight distribution management. Top **25**, 413-433 (2017). doi: 10.1007/s11750-017-0456-1.
4. Beardwood, J., Halton, J.H., Hammersley, J.M.: The shortest path through many points. Mathematical Proceedings of the Cambridge Philosophical Society **55**(4), 299-327 (1959). doi: /10.1017/S0305004100034095.
5. Vig, V., Palekar, U.S.: On estimating the distribution of optimal traveling salesman tour lengths using heuristics. European Journal of Operational Research **186**(1), 111–119 (2008). doi: /10.1016/j.ejor.2006.12.066
6. Johnson, D.S., McGeoch, L.A., Rothberg, E.E.: Asymptotic experimental analysis for the Held-Karp traveling salesman bound: In the Proceedings of the 7th Annual ACM-SIAM Symposium on Discrete Algorithms, pp. 341-350, Atlanta, Georgia (1996).
7. Deza, E., Deza, M. M.: Encyclopedia of distances. Springer-Verlag Berlin, Heidelberg (2014).
8. Koshizuka, T., Kobayashi, J.: On the Relation between Road Distance and Euclidean Distance, Journal of the City Planning Institute of Japan **18**, 43-48 (1983). doi: /10.11361/journalcpij.18.43 (in Japanese).
9. Kurita, O.: Theory of Road Patterns for A Circular Disk City, Journal of the City Planning Institute of Japan **36**, 859-864 (2001). doi: /10.11361/journalcpij.36.859 (in Japanese).
10. Çavdar, B., Sokol, J.: A distribution-free TSP tour length estimation model for random graphs. European Journal of Operational Research **243**(2), 588-598 (2015). doi: /10.1016/j.ejor.2014.12.020.
11. Holguín-Veras, J., Patil, G.R.: Observed trip chain behavior of commercial vehicles. Transportation Research Record **1906**(1), 74–80 (2005). doi: /10.1177/0361198105190600109
12. Choi, Y., Schonfeld, P.M.: Review of length approximations for tours with few stops. Transportation Research Record **2676**(3), 201–213 (2022). doi: /10.1177/03611981211049433

13. Chien, T.W.: Operational estimators for the length of a traveling salesman tour. Computers & Operations Research **19**(6), 469–478 (1992). doi: /10.1016/0305-0548(92)90002-M
14. Helsgaun, K.: An effective implementation of the Lin–Kernighan traveling salesman heuristic. European Journal of Operational Research **126**(1), 106–130 (2000). doi: /10.1016/S0377-2217(99)00284-2
15. Johnson, D.S., McGeoch, L.A.: Experimental analysis of heuristic for the STSP. In: Gutin, G., Punnen, A.P. (eds.) The Traveling Salesman Problems and Its Variations. Kluwer Acdemic Publishers, pp. 369–443, Springer New York, New York (2002).