

A time-varying p -median model for location-allocation analysis

Shahin Sharifi Noorian
Delft University of Technology
Van Mourik Broekmanweg 6
2628 XE Delft
Delft, the Netherlands
s.sharifinoorian@tudelft.nl

Achilleas Psyllidis
Delft University of Technology
Van Mourik Broekmanweg 6
2628 XE Delft
Delft, the Netherlands
a.psyllidis@tudelft.nl

Alessandro Bozzon
Delft University of Technology
Van Mourik Broekmanweg 6
2628 XE Delft
Delft, the Netherlands
a.bozzon@tudelft.nl

Abstract

Location models have traditionally played an important role in suggesting sites for the placement of facilities, so that efficient service delivery is ensured. A common formulation of several location models is associated with the p -median problem, which aims to minimize the travel distance between support facilities and demand in a region. However, the influence of external conditions, such as traffic, on travel time is largely ignored. In this paper, we present a time-varying approach to the classical p -median problem, which accounts for fluctuations in travel cost distance at different time intervals. Using Google Traffic and Foursquare data to respectively retrieve traffic information and estimate demand in a region, and by employing an adaptive genetic algorithm in a planning problem application in the Netherlands, we show that our proposed model outperforms the classical p -median formulation, in providing more travel efficient service of demand nodes. Moreover, we achieve better placement of support facilities across major street arteries. The paper concludes with a discussion of associated uncertainties that are important to be recognized prior to viewing the modeling results as suggestions for implementation in planning and policy making.

Keywords: location-allocation; p -median; genetic algorithm; time-varying location model; social data; traffic.

1 Introduction

The siting of product supply facilities plays an important role in the efficiency of service delivery for commercial activities. Distance from and travel time to a demand site are, therefore, critical factors in the evaluation of a service location. In suggesting optimal sites for the placement of services, various types of location models have been developed and used over the past five decades (Church, 1999). Many of these models have approached the location problem as a p -median problem (Daskin & Maass, 2015). In this, a number of discrete facilities p are located in such a way that the weighted sum of travel costs from support facilities to demand in a region is minimized.

Given that location models are abstractions of reality, a number of uncertainties associated with the model parameters is involved. In the case of the p -median problem, one such uncertainty is associated with the definition of the *travel cost*. In the majority of approaches to the p -median problem, the travel cost is calculated using the Euclidean or Manhattan (network-based) distance between a support facility and a demand area. This implies that travel cost is a static parameter, ignoring the influence that external conditions, such as traffic, could have on its estimation. In the real world, however, travel cost is a time-varying attribute. The incorporation of travel cost information has been hampered by a lack of high-resolution data about real-time traffic conditions. Yet, the recent proliferation and availability of new forms of data about traffic (e.g. Google Traffic) make possible the extraction of fine-grained traffic information over various time intervals (Psyllidis, 2015).

In this paper, we introduce a novel formulation of the classical p -median problem that explicitly incorporates time

and traffic conditions in the estimation of the travel cost parameter. We use Google Traffic data to retrieve traffic information at different times of the day and on various days of the week; and Foursquare data to estimate the demand for service in a region. The distance between the demand areas and the support facilities relies on the street network, as extracted from Google Maps. The individual demand sites (e.g. restaurants) retrieved from Foursquare are aggregated at the municipal level to reflect the demand in a region (e.g. municipality). In approximating near optimal solutions to the proposed model, a metaheuristic approach using a genetic algorithm with self-adaptive and fast convergence mutation is presented. Results show that the proposed model provides (on average) 6.5% more travel efficient service of demand in a region, compared to the classical p -median. Moreover, the resulting locations of facility sites are placed across major street arteries, thereby facilitating service delivery to demand.

2 Related Work

The p -median problem is a well-known NP-hard location problem which was originally described by (Hakimi, 1964), and formulated mathematically by (ReVelle and Swain, 1970). Existing literature on location modeling has contributed a rich variety of p -median models with different objective function specifications and associated constraints. Despite the wide range of possible modeling variants, the travel distance between facilities and demand has primarily been considered constant. Yet, in reality, travel distances between locations can vary largely due to, for instance, traffic conditions.

There exist a few studies that have investigated dynamic location-allocation models. Bloxham and Church (1991)

introduced a p -median scheduling model, in which the facilities are located based on time of operation and demand is allocated only to open facilities. Louveaux (1986) has formulated a deterministic model of the incapacitated facility location problem, known as the simple plant location problem, in which demands, variable production and transportation costs as well as selling prices can be changed over time. Drezner (1995) developed a dynamic p -median model to find the best location to build a new facility in addition to existing ones. Chukwusa (2014) explored the impact of spatiotemporal variations in demand and suggested a trend-weighted location-allocation model to solve the p -median problem when the demand is changing over time.

3 Time-varying p -median model

In this section, we present the mathematical formulation of the proposed time-varying p -median problem. We consider the following notation:

- i = demand areas ($1, 2, \dots, n$)
- j = candidate facility sites ($1, 2, \dots, m$)
- k = potential departure time slots for a facility to serve a demand area ($1, 2, \dots, l$)
- p = number of facilities to be located
- c_{ijk} = travel time from demand area i to facility j at departure time k
- a_i = amount of demand in area i
- X_{ij} = demand i is served by facility j : 1 , otherwise: 0
- Y_j = facility at site j is located: 1 , otherwise: 0
- T_{ijk} = demand i is served by facility j at departure time k : 1 , otherwise: 0

And formulate the proposed model as follows:

$$\text{Min} \sum_{i=1}^n \sum_{j=1}^m \sum_{k=1}^l a_i c_{ijk} X_{ij} T_{ijk} \quad (1)$$

$$\text{Subject to:} \sum_{j=1}^m X_{ij} = 1, \text{ for each } i=1, \dots, n \quad (2)$$

$$X_{ij} \leq Y_j, \text{ for each } i=1, \dots, n \text{ and } j=1, \dots, m \quad (3)$$

$$\sum_{k=1}^l T_{ijk} = 1, \text{ for each } i=1, \dots, n \text{ and } j=1, \dots, m \quad (4)$$

$$\sum_{j=1}^m Y_j = p \quad (5)$$

$$X_{ij} \in \{0, 1\}, \text{ for each } i=1, \dots, n \text{ and } j=1, \dots, m \quad (6)$$

$$Y_j \in \{0, 1\}, \text{ for each } j=1, \dots, m \quad (7)$$

$$T_{ijk} \in \{0, 1\}, \text{ for each } i=1, \dots, n \text{ and } j=1, \dots, m \text{ and } k=1, \dots, l \quad (8)$$

The objective function (1) minimizes the total weighted travel cost between a support facility and demand sites in a region. The distance between facilities and demand sites is calculated using the street network. T_{ijk} represents the allocation decision about which demand area is served by what support facilities at a given departure time k , while accounting for the traffic conditions on a given day of the week and at a given time of the day.

Constraints (2-4) and the objective function (1) stipulate that each demand i will be allocated to their closest supply facility

j at a given departure time k . Constraint (5) indicates that p sites will be placed. Constraints (6-8) dictate that the location and allocation variables should be integer and binary, i.e. that each demand area should be allocated to exactly one support facility at a given departure time.

Given that the model (1) is non-linear, it is difficult to provide a deterministic solution to it. Therefore, a metaheuristic method is proposed for solving this problem, based on an adaptive genetic algorithm.

4 Genetic algorithm with self-adaptive and fast convergence mutation

Given the large number of possible solutions that are associated with non-linear problems such as the p -median, we adopt a metaheuristic approach using a genetic algorithm. GAs have proven to be efficient and effective in providing solutions to non-linear problems with large instances (Li 2011; Xiao, 2005). An additional advantage of GAs is associated with their ability to find a population of good solutions, while maintaining the diversity among them.

If f_i represents a facility located at site i , then the index of all located facilities could be encoded as a chromosome of the GA with length p (i.e. the number of facilities to be located), in the form $[f_1, f_2, \dots, f_p]$. Algorithm 1 computes the fitness function of the proposed GA, which corresponds to the objective function (1) of the time-varying p -median problem. The fitness function ensures that all demand areas are allocated by the nearest facility at a specific time slot and with the shortest travel cost distance. It further guarantees that all of the defined constraints (2 – 8) are satisfied.

Algorithm 1 GetFitness()

Inputs:

$Chromosome = [facility_1, \dots, facility_N]$

$D = [d_1, d_2, \dots, d_M]$

$T = [t_1, t_2, \dots, t_L]$

Initialize:

$TotalCost \leftarrow 0$

for $d \in D$ **do**

$MinimumTravelTime \leftarrow \infty$

for $facility \in Chromosome$ **do**

for $t \in T$ **do**

$TravelTime \leftarrow CalculateTravelTime(facility, d, t)$

if $TravelTime \leq MinimumTravelTime$ **then**

$MinimumTravelTime \leftarrow TravelTime$

end if

end for

end for

$TotalCost \leftarrow TotalCost + MinimumTravelTime$

end for

return $TotalCost$

GAs are prone to premature convergence to local optima. We address the problem by adjusting the mutation rate (P_m) while the algorithm explores the search space, as also proposed in (Galaviz and Xuri, 1995). To avoid the generation of invalid solutions and to improve the performance of the GA, we use a greedy approach to the mutation process, which mutates the offspring only if the mutated solution gains lower

fitness value. The refined mutation process which incorporates the aforementioned greedy approach is outlined in Algorithm 2.

Algorithm 2 GreedyMutation()

Inputs:
 P_m : mutation rate
 C : chromosome to be mutated
 $H \leftarrow \text{GetFitness}(C)$
 $C' \leftarrow \text{Mutate}(C, P_m)$
 $H' \leftarrow \text{GetFitness}(C')$
if $H > H'$ **then**
 return C'
else
 return C
end if

Population, elitism, mutation-rate and cross-over rate are regarded as hyper parameters, and therefore can be found via grid search ($N \in \{50, 100, 150, 200\}$; elitism $E \in \{5, 10, 15, 20, 30\}$; mutation rate $P_m \in \{0.0005, 0.001, 0.005, 0.01, 0.05\}$; and cross-over rate $P_c \in \{0.5, 0.6, 0.7, 0.8, 0.9\}$). We set $N = 100$; $E = 20\%$; $P_m = 0.005$; and $P_c = 0.7$. We used a stagnation-based termination criterion; following (Alp et al., 2003), we terminate the algorithm after $n\sqrt{p}$ generations, where p represents the number of facilities to be located, and n is the number of demand areas. The proposed GA algorithm is available on GitHub¹.

5 Datasets

The application of our proposed time-varying p -median problem intends to identify the best siting configuration of supply facilities and allocation schemes of demand for service, in a context where time plays an important role. We focus on fish markets (i.e. supply facilities) and restaurants (i.e. demand) and selected the Netherlands as reference area. Travel time plays an important role in securing the freshness of the supplied products to the various demand areas. Using the Foursquare API², we retrieve 60,000 restaurants in the Netherlands, which we aggregate into 387 municipal regions (the islands in the Wadden region were excluded, as there is no street connection). The number of restaurants in each municipality represents the corresponding demand in that area. Without loss of generality, we use the centroid of each municipal region as the candidate location to site the various supply facilities, and limit the reference time slots to *Monday 8am*, *Monday 12pm*, and *Monday 6pm*. We use the Google Maps Distance Matrix API³ to extract the duration and distance values between the centroids of the municipal regions. The resulting travel time matrix comprises 154,056 records.

¹ <https://github.com/shahinsharifi/AGILE2018>

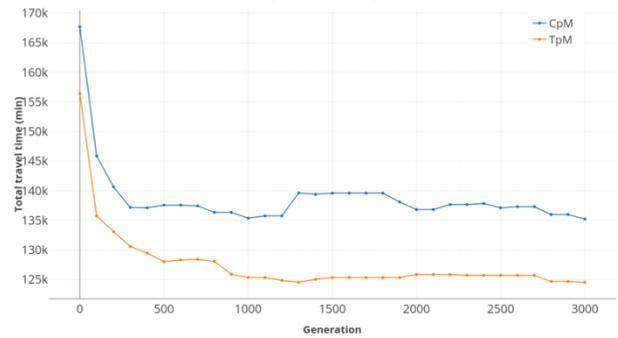
² <https://developer.foursquare.com/places-api>

³ <https://developers.google.com/maps/documentation/distance-matrix/>

6 Results

Code has been implemented in Java 8, and experiments were carried out on a personal computer with Intel Core i7-4770 3.40 GHz (Quad core CPU), 8 Gigabyte RAM memory, Hard Drive with 7200 RPM, under Linux kernel 4.13.0-32-generic. We evaluated different siting configurations, with p varying from 5 to 40. We carried out different experiments, using both the classical p -median formulation (CpM) — in which, travel cost distance is considered a static parameter — our proposed time-varying model (TpM), and compared the corresponding results. We applied the GA described in Section 4, 5 times to each one of the models, and averaged the results. The best identified solutions for $p = 20$ are shown in Figure 2. By performing the GA algorithm for both TpM and CpM , using the same configuration, and with $p = 20$, we find that the quality of solution in terms of the minimization of travel cost is improved in TpM , as depicted in Figure 1. The corresponding siting configurations for both models is illustrated in Figure 2. The location of the sited facilities is represented with a house-like symbol, whereas the demand areas are classified by means of different colors based on the index of corresponding facility.

Figure 1: Total travel time calculation with the conventional p -median model (CpM) and the Time-varying p -median model (TpM), with $p = 20$.



Results for the planning application for the different p instances are presented in Table 1. In all instances, our proposed model appears to outperform the classical p -median approach, in terms of identifying facility locations that can serve faster the corresponding demand. However, the computational effort needed to find near-optimal solutions to our model is on average 20-30% higher than the one required for the classical problem. This is due to the additional operations required to find the best time slot for an allocated demand area.

Table 1: Comparison of total cost (travel time) and execution time for CpM and TpM , at varying values of p .

P	CpM		TpM	
	Total cost (min.)	Execution time (sec.)	Total cost (min.)	Execution time (sec.)
5	260,859	25	246,652	37
10	203,139	30	188,358	112
20	163,065	198	158,593	222
40	135,224	515	124,794	534

Figure 2: Siting configuration with (a) the conventional (CpM) and (b) time-varying (TpM) p -median models, with $p=20$, and (c) overlay map of the two models (fill: TpM , outline: CpM).

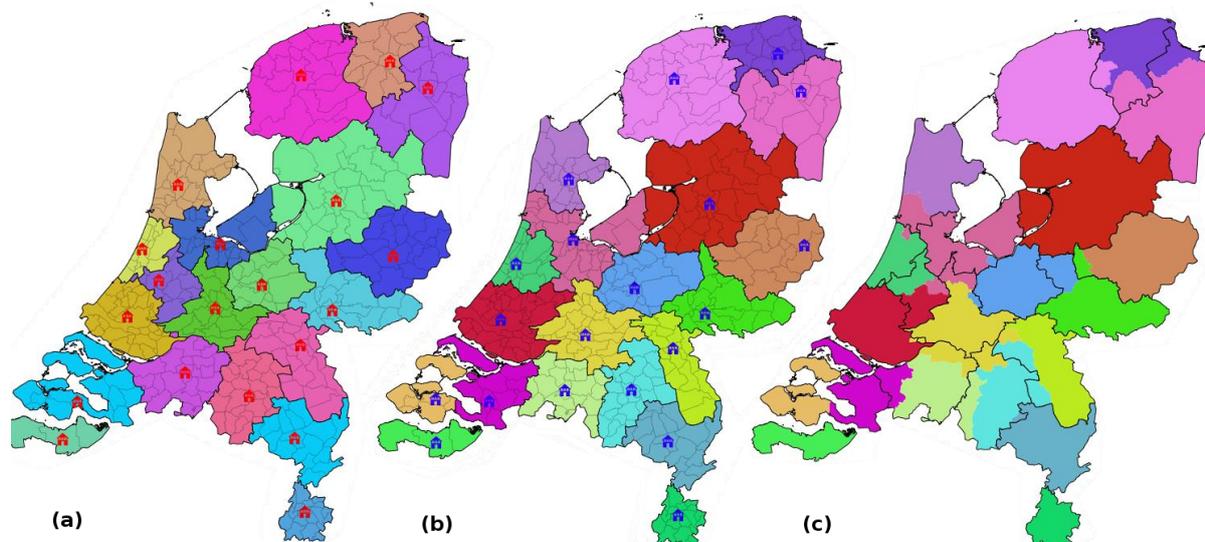


Table 2 shows how differences in time slots affect demand allocations, as serviced by 10 supply facilities. Results show that all demand areas are allocated by 10 facilities and for each facility-demand relationship, a certain time slot is assigned – thus fulfilling constraints (5), (6), (7), and (8) of Section 3.

Table 2: Comparison of demand allocations at different time slots and $p = 10$.

Facility	Demand Areas	Time slots		
		Monday 8am	Monday 12pm	Monday 18pm
1	61	5	31	25
2	23	2	1	20
3	48	6	19	23
4	29	1	1	27
5	23	0	6	17
6	29	1	0	28
7	60	0	57	3
8	27	0	1	26
9	54	2	46	7
10	32	2	5	25
	387	19	167	201

The amount of allocated demand areas varies between different time slots. For example, 31 out of 61 demand areas serviced by *Facility 1* should be served during the second time slot due to shorter travel time. On the other hand, there is no demand serviced by *Facility 5* at the first time slot; a result that could be explained by heavier traffic conditions on the way between *Facility 5* and the demand areas allocated to it on *Monday at 8am*.

In order to prove that the facilities have been sited at places with good accessibility, we have overlaid the result of our proposed model on the road network of the Netherlands. Figure 3 clearly shows that the location of facilities and allocated demand areas generated by TpM are placed across major road arteries.

Figure 3: Overlay of Dutch road network, sited facilities, and allocated demand areas.



7 Discussion and Conclusions

This paper has proposed a time-varying formulation of the classical p -median problem that accounts for the influence of traffic conditions on the estimation of the travel costs between support facilities and demand areas.

This has been possible with the recent proliferation of new forms of data sources, such as Google Maps, Google Traffic, and Foursquare, which provide fine-grained information about how distances between locations are affected by traffic conditions at different time intervals, as well as about the estimated demand for service.

The results presented in Table 1 and Table 2 indicate that our proposed model outperforms the classical p -median problem formulation, in which travel cost is based on the Euclidean distance, and external conditions such as traffic are not considered.

Our model provides more efficient siting configurations and allocation schemes, in terms of the time required for a service to supply the demand in a region. In addition, as shown in Figure 3, the siting of support facilities is better integrated in the street network. In order to emphasize the influence of traffic conditions on the location-allocation analysis, we have intentionally used Monday traffic data, given that high peaks usually occur on that day of the week. However, we have further tested our proposed model on data pertaining to the rest of the weekdays (not presented here due to space limitations), and it consistently outperforms the conventional p -median model.

Although the aforementioned findings are promising, uncertainties associated with the data sources and the modeling analysis could impact the modeling results. Data collected from emerging sources, such as Google Traffic and Foursquare, are prone to biases associated with representativeness and measurement error. For instance, Foursquare does not capture the entire set of available functions (in this case, restaurants) in a region, which could lead to false estimates of the demand for service. In addition, there is spatial variation in quality, meaning that in certain locations the data about various functions could be more detailed and comprehensive than in other parts of a region. Similarly, data about traffic conditions may contain missing values, as well as be prone to measurement error, given that they usually derive from GPS sensors embedded in mobile phones. In addition to data uncertainties, modeling results could be influenced by the mathematical formulation of the problem and the method used for solving it. In this paper, we have employed a GA approach to the modeling analysis. In future work, we plan to examine alternative methods, such as tabu search and simulated annealing, so as to evaluate the set of solutions associated with the proposed model. Data and modeling analysis uncertainties, as previously described, are important to be recognized prior to viewing the modeling results as suggestions for implementation in planning and policy making.

Despite these uncertainties, our proposed model holds promise as an extension to the classical p -median problem, by effectively accounting for traffic conditions when allocating demand to facility services.

References

- Alp, O., Erkut, E., and Drezner, Z. (2003) An efficient genetic algorithm for the p -median problem. *Annals of Operations research*, 122(1-4), 21–42.
- Bloxham, C.A. and Church, R.L. (1991) The p -median scheduling and location problem. *Papers in Regional Science*, 70(1), 21–35.
- Cantu-Paz, E. and Goldberg, D. E. (2000) Efficient parallel genetic algorithms: Theory and practice. *Computer Methods in Applied Mechanics and Engineering*, 186(2-4), 221–238.
- Chukwusa, E. (2014) *The impact of alternative distance measures and temporal variation in demand on location-allocation decisions*. PhD dissertation, Department of Geography, University of Leicester, UK.
- Church, R. L. (1999) Location modelling and GIS. In: P. Longley, M. Goodchild, D. Maguire, and D. Rhind (eds.), *Geographical Information Systems* (second ed.). New York: John Wiley, pp. 293–303.
- Daskin, M. S. and Maass, K. L. (2015) The p -median problem. In: G. Laporte et al. (eds.) *Location Science*. Springer International Publishing Switzerland, pp. 21–45.
- Drezner, Z. (1995) Dynamic facility location: The progressive p -median problem. *Location Science*, 3(1), 1–7.
- Galaviz, J. and Xuri, A. (1996) A self-adaptive genetic algorithm for function optimization. In: *Proceedings of IEEE ISAI/IFIS Collaboration in Intelligent Systems Technologies*, Mexico.
- Hakimi, S. L. (1964) Optimum location for switching centers and the absolute centers in a graph. *Operations Research*, 12(1), 450–459.
- Li, X., Xiao, N., Claramunt, C., and Lin, H. (2011) Initialization strategies to enhancing the performance of genetic algorithms for the p -median problem. *Computers & Industrial Engineering*, 61(4), 1024–1034.
- Louveaux, F. V. (1986) Discrete stochastic location models. *Annals of Operations research*, 6(2), 21–34.
- Psyllidis A., Bozzon A., Bocconi S., Titos Bolivar C. (2015) A Platform for Urban Analytics and Semantic Data Integration in City Planning. In *Computer-Aided Architectural Design Futures. The Next City - New Technologies and the Future of the Built Environment. CAAD Futures 2015*. Communications in Computer and Information Science, vol 527. Springer, Berlin, Heidelberg
- ReVelle, C. S. and Swain, R. (1970) Central facilities location. *Geographical Analysis*, 2(1), 30–42.
- Xiao, N. (2005) Geographic optimization using evolutionary algorithms. In *Proceedings of the 8th International Conference on GeoComputation*, Ann Arbor, Michigan, USA.