Transactions in GIS, 2008, 12(3): 323–339

Research Article

A Clustering-Based Approach to the Capacitated Facility Location Problem[1]

Ke Liao
Department of Geography
University of South Carolina

Diansheng Guo
Department of Geography
University of South Carolina

Abstract

This research develops a clustering-based location-allocation method to the Capacitated Facility Location Problem (CFLP), which provides an approximate optimal solution to determine the location and coverage of a set of facilities to serve the demands of a large number of locations. The allocation is constrained by facility capacities – different facilities may have different capacities and the overall capacity may be inadequate to satisfy the total demands. This research transforms this special location-allocation problem into a clustering model. The proposed approach has two parts: (1) the allocation of demands to facilities considering capacity constraints while minimizing the cost; and (2) the iterative optimization of facility locations using an adapted K-means clustering method. The quality of a location-allocation solution is measured using an objective function, which is the demand-weighted distance from demand locations to their assigned facilities. The clustering-based method is evaluated against an adapted Genetic Algorithm (GA) alternative, which integrates the allocation component as described above but uses GA operations to search for 'optimal' facility locations. Experiments and evaluations are carried out with various data sets (including both synthetic and real data).

1 Introduction

The location of facilities and allocation of demands to facilities is a spatial decision problem that has many applications, e.g. retail site selection (Goodchild 1984), industrial logistics (Ceselli and Righini 2005), marketing analysis (Díaz and Fernández 2005), transportation planning (Horner and O'Kelly 2005), power distribution system design (Haghifam and Shhabi 2002), telecommunication network design (Kalvenes and Kennington 2005; Kratica et al. 2005), and emergency response system management (Ehrgott 2002). A solution to a general location-allocation problem should decide both

Address for correspondence: Ke Liao, Department of Geography, University of South Carolina, 709 Bull Street, Columbia, SC 29208, USA. E-mail: liao4@mailbox.sc.edu

the location and the coverage of each facility. The solution is typically evaluated with an overall cost measure, for example, the total distance between facilities and demand locations that they serve.

Location-allocation problems have been studied extensively, with various constraints and considerations in real world applications. The main constraint addressed in this article is the capacity limitation of service suppliers and the adequacy of their total capacity, which should be considered to ensure acceptable level of service and spatial equity (Nauss 1978, Murray and Gerrard 1997, Chudak and Williamson 2005). For example, emergency planners may need to locate and establish a given number of shelters in a hurricane-prone zone. A capacity limitation is imposed on each shelter to avoid overutilization and underutilization of a shelter. Moreover, due to limited resources, the total capacity of all shelters may be insufficient and some areas will be left uncovered. The emergency shelters should be optimally located so that the covered population (to be decided by the solution) can access the shelters with minimum cost.

This article presents a clustering-based approach to solve such capacity constrained location-allocation problems. The approach is similar to the K-means clustering method, which is adapted to accommodate capacity constraints (i.e. cluster size constraints). The clustering-based approach also incorporates several strategies to avoid local optima. The proposed approach is evaluated against a hybrid method based on a genetic algorithm (only for searching facility locations) and the allocation component as used in the above clustering-based approach. The remainder of the paper is organized as follows. The next section reviews relevant research work. Section 3 introduces the clustering-based method and its GA alternative. Section 4 reports the experiment result with different distribution of demands. Section 5 provides a summary of the findings and a brief discussion of future work.

2 Related work

Most capacitated location-allocation models implicitly assume adequate capacity (Nauss 1978, Chudak and Williamson 2005). This research concerns both sufficiently capacitated and insufficiently capacitated facility location problems. Uncovered locations are possible if the total capacity of facilities is not sufficient to cover all locations. In this research, one demand location can only be assigned to one facility, i.e. fractional assignment is not allowed. Let $\mathbf{I} = \{1 \ldots n\}$ be all the demand locations and $\mathbf{J} = \{1 \ldots m\}$ be all the facilities to serve the demand locations. The overall quality of a solution $\mathbf{S}$ is measured by $\mathbf{V(s)}$, which is defined as the demand-weighted distance from demand locations to their assigned facilities. Such capacitated location-allocation problems (CFLP) can be formulated as follows.

1.1 $$\min V(S) = \sum_{j=1}^{m}\sum_{i=1}^{n} x_{ij} p_i d_{ij},\ \forall\ i \in I, j \in J$$

subject to:

1.2 $$x_{ij} = \{0, 1\},\ \forall\ i \in I, j \in J$$

1.3 $$\sum_{j=1}^{m} x_{ij} = \{0, 1\},\ \forall\ i \in I, j \in J$$

1.4 $$\sum_{i=1}^{n} x_{ij}\ a_i \leq b_j,\ \forall\ i \in I,\ j \in J$$

where

$$p_i = \frac{a_i}{\sum_{j=1}^{m}\sum_{i=1}^{n} x_{ij}\ a_i},\ \forall\ i \in I,\ j \in J$$

$$d_{ij} = \left[\sum_{k=1}^{2}(s_{ik} - y_{jk})^2\right]^{1/2},\ \forall\ i \in I,\ j \in J$$

$$\| I \| = n,\ \| J \| = m.$$

The binary variable $x_{ij} = 1$ indicates that a demand location i is assigned to a facility j. One location can only be assigned to one facility (1.3). The capacity constraint requires that the demands assigned to a facility be less than or equal to its given capacity (1.4). Since there is no limitation on the overall capacity, the total capacity can be sufficient or insufficient to cover all the demands. The weight of each assignment (p_i) is calculated as the percentage of demands at location i of the total demands assigned to that facility. The cost metric (d_{ij}) represents the access cost between a demand location and a facility. It can be a Euclidean distance (as used in this research) or other measures such as travel time. V(s) is affected by the location of facilities and the assignment of demands. The goal of CFLP is to find optimal (or near-optimal) locations of a finite set of facilities and decide the coverage of each of them under capacity constraints and assignment constraints.

Location-allocation problems with capacity constraints have many variants across different application contexts, including the regionally constrained p-median problem (RCPMP) (Murray and Gerrard 1997), the capacitated p-median problem (CPMP), the capacitated centered clustering problem (CCCP) (Negreiros and Palhano 2006), the capacitated single allocation hub location problem (Ernst and Krishnamoorthy 1999), the single source capacitated plant location problem (Díaz and Fernández 2002), among many others. They are different from each other in terms of different constraints on demand assignments and/or facility locations. For example, CPMP requires that facility locations be a subset of the demand point locations (Díaz and Fernández 2005). RCPMP requires some facilities to be placed in given regions (Murray and Gerrard 1997). The research reported in this article allows a facility location in the solution to be different from demand locations. It also forces the solution to locate exactly the given number of facilities.

The allocation of capacities is often addressed as a subproblem in location-allocation analysis. With facility locations fixed, the goal becomes finding an allocation with minimum access cost. This objective can be accomplished by using a flow network model, where supplier locations and demand locations are defined as nodes and edges between nodes are associated with a cost value and a capacity level. Various algorithms are developed to find optimal solutions to minimum-cost flow network models with various objective functions (Sedeño-Noda and González-Martín 2001, Bertsimas and Sim 2003). One typical application of a flow network model is to solve transportation and transshipment problems (Ahuja et al. 1993).

Capacitated location-allocation problems can be structured as a Linear Programming (LP) problem with a linear combination of solution variables (Vinod 1969, ReVelle and

Swain 1970). However, capacitated location-allocation problems are *NP*-complete (Garey and Johnson 1979). The time cost of a deterministic approach will increase exponentially and make it impractical to process large location-allocation problems. Therefore, substantial research work has been carried out to develop heuristics to obtain good approximations of the optimal solution (França et al. 1999, Wu et al. 2006). Heuristics can be integrated with LP models or applied as standalone methods, such as branch-and-bound (Marín and Pelegrín 1997), simulated annealing (SA) (Ernst and Krishnamoorthy 1999), bionomic metaheuristic algorithm (Maniezzo et al. 1998), adaptive tabu search (França et al. 1999), set partitioning (Baldacci et al. 2002), column generation (Lorena and Senne 2004), and scatter search (Díaz and Fernández 2005, Scheuerer and Wendolsky 2006). Earlier heuristics proposed for general location-allocation problems include TORNQVIST (Kohler 1973) and an improved vertex substitution method (Densham and Rushton 1992).

The research reported here uses a clustering strategy (which is a special heuristic approach) to formulate solutions for location-allocation problems with capacity constraints. Clustering analysis is one of the most commonly used approaches in data analysis and has been applied in many application domains such as pattern discovery, document retrieval, image segmentation, among many others. Clustering methods do not depend on prior knowledge and can discover natural groupings of data items (Jain and Dubes 1988, Jain et al. 1999, Han et al. 2001; Guo et al. 2003). Therefore, when used in a location-allocation context, clustering methods have the capability to adapt to the distribution of demands and thus facilitate the search for an approximate optimal solution. Mulvey and Beck (1984) presented a clustering-based heuristic (CAPCLUST), whose performance is very close to that of an LP-based approach. The implicit connection between this clustering concept and location-allocation problems is also suggested in an earlier work (Vinod 1969).

Both the CAPCLUST method and our proposed method are adapted from the K-means clustering algorithm. K-means is a distance-based partitioning clustering approach that partitions a set of data items into clusters while ensuring a low internal dissimilarity or distance. It assumes that the number of clusters (k) is known. The K-means algorithm consists of three steps (Jain et al. 1999): (1) randomly choosing k cluster centers within the data space; (2) assigning each data item to the closest cluster center; and (3) recalculating the cluster centers using the points assigned to each cluster. Steps 2 and 3 are then repeated until the result converges. To adapt it to a location-allocation problem, Step 2 can be used to allocate demands and Step 3 can be used to optimize facility locations.

The method developed in this research is different from the CAPCLUST method in several aspects: (1) the consideration of both insufficient and sufficient capacity constraints (CAPCLUST assumes overall excess capacity); (2) the use of a facility-based swap strategy (see Section 3), which is more efficient and has a greater potential in finding globally optimal facility locations than the demands-based switch performed in CAPCLLUST; and (3) the incorporation of several other improvements to avoid local optima (see Section 3). With a clustering-based framework for location-allocation problems, the optimization of allocation becomes easier (Step 2, see above) and the focus therefore is more on the optimization of facility locations.

3 A K-means Clustering-Based Approach to CFLP

This section describes a clustering-based method to solve CFLP. The method is adapted from the K-means clustering algorithm (Figure 1a). A capacity constraint is considered

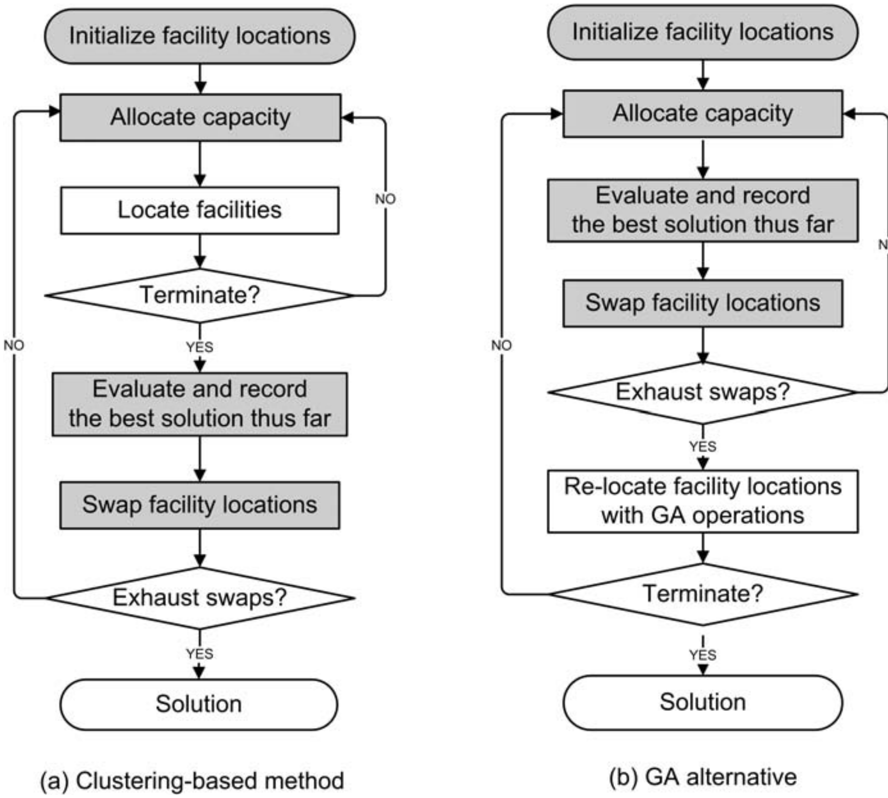

Figure 1 Schematic showing clustering-based approach and a GA alternative

during the allocation of capacities, i.e. the assignment of demand points to facilities is subject to the limitation of facility capacities. Starting with an initial set of facility locations, the capacity-constrained assignment of demand points and the optimization of facility locations are iterated, similar to Steps 2 and 3 in K-means clustering (see Section 2 and Figure 1a). Once the iteration terminates (i.e. an allocation-location solution is obtained), the determined facility locations are swapped among facilities and the iterative allocation-location is carried out again after each facility-location swap to obtain a new solution. This swap step is to escape local optima and is not needed if all facilities have the same capacity. A V(s) value is calculated for each solution and the one with the smallest V(s) value will be selected as the final solution.

The overall capacity can be either sufficient or insufficient to accommodate all demands. The clustering-based method has two slightly different procedures to deal with these two different situations (i.e. sufficient or insufficient total capacity). The procedure for insufficiently capacitated problems is different in two aspects: (1) it includes an additional step to avoid local optima; and (2) it uses a different priority measure in assigning demand points to facilities. Sections 3.1 and 3.2 introduce these two different procedures for sufficient capacity and insufficient capacity, respectively.

A hybrid genetic algorithm (GA) method (Figure 1b) is also proposed as an alternative for the clustering-based method. This GA-based method uses the same allocation component as used in the clustering-based method but replaces the facility location optimization component with an adapted GA-based approach. The common steps in the clustering method and its GA variant are shaded in Figure 1. Since the genetic algorithm is a global searching heuristic, it is used to assess the performance of the clustering-based method in finding approximations of globally optimal location-allocation solution. Section 3.3 focuses on this hybrid GA alternative.

3.1 Sufficient Capacity

When the overall capacity of facilities is adequate to meet the overall demand, the proposed clustering-based method consists of the following steps (see Figure 1a):

(Step 1) **FacilityInitialization**: Initialize facility locations randomly. Facility locations are not limited to be among the demand locations. Increase the capacity of each facility proportionally such that the overall capacity is larger (e.g. by 20%) than the overall demand.

(Step 2) **AllocationAndLocation**: Run an iterative allocation-location process. Each iteration involves two operations: (a) assigning demand locations to their closest unfulfilled facilities; and (b) recalculating each facility location as the centroid of its service area using the results of the first stage. Repeat (a) and (b) until facility locations converge or the maximum of iterations is reached.

(Step 3) **FacilityLocationSwap** (optional – only needed when facility capacities are different from each other): Swap those final locations determined in Step 2 exhaustively among facilities and use the original capacity values (not the increased values in Step 1). This step includes three operations: (a) swap the facility locations among facilities; (b) run the allocation-location iteration after each capacity swap; and (c) evaluate the result by V(s) (and record the result if it is better than the previously recorded best solution). Repeat (a) and (b) until the facility locations are coupled with the actual capacities in all possible ways.

(Step 4) **OptimalSolution**: Choose the final solution with the minimum V(s) among all recorded solutions.

The first two steps (i.e. *FacilityInitialization* and *AllocationAndLocation*) are designed to find a good configuration of facility locations given their capacities. In this article, "capacity allocation" and "demand assignment" are used interchangeably, which is to assign demands to facilities. In order to prevent local optima, the actual capacities are enlarged proportionally at the beginning (Step 1). With such a relaxed (i.e. increased) capacity, facility locations can better adapt to the distribution of demand locations.

The *AllocationAndLocation* iteration is based on Step 2 and Step 3 in the K-means algorithm (see Section 2) with two modifications. First, capacity constraint is considered by checking whether the facility is full or not before assigning a demand location to it. Secondly, similar to CAPCLUST (Mulvey and Beck 1984), demand points are assigned in a descending order of a priority value to attain a low V(s). The priority value is calculated for each demand point as the absolute difference in the distances to its first and second closest facilities that are not fully loaded. A demand point with a higher priority value is more competitive in being assigned to its nearest facility. The allocation-location procedure (step 2) can achieve a local optimal location solution. The algorithm for this step is summarized in Table 1.

Table 1 The allocation-location algorithm used to achieve a local optimal location solution

Set Δ to 10,000 (or any large number)
Set MaxIteration = 1000 (or any large number)
Set IterationCount = 0
Set ε = a small number as the convergence criterion
While (Δ > ε) and (IterationCount < MaxIteration)
Increase IterationCount by 1
Compute a priority value and find the nearest unfulfilled facility for each demand point
Sort the demand points with their priority values in a descending order
  For each demand point in the sorted list (following the sorted order)
    If the demand point is not assigned yet
    Assign the demand point to its nearest unfulfilled facility
    If this facility becomes full after the assignment
    Recalculate priority value (and nearest unfulfilled facility) for each remaining demand point
    Sort the remaining demand points (note: now the sorted list contains only the remaining points)
  End
  For each facility
    Compute facility location as the centroid of demand points that the facility covers
    Compute Δ (i.e., difference between the new and previous facility locations)
  End
  End while

The *FacilityLocationSwap* step (i.e. Step 3) is to evaluate all possible mappings between the derived facility locations in step 2 and the facilities (whose capacities are different). For example, if there are three facilities (A, B, C) with different capacities (e.g. 50, 100, 150), Step 2 may place facilities A, B, and C at locations L_1, L_2, and L_3, respectively. One possible swap would position facility A at L_2, facility B at L_3, and facility C at L_1. If all facilities have the same capacity, this *FacilityLocationSwap* step is not needed and skipped. After each facility-location swap, step 2 is repeated to re-configure the facility locations and obtain a new allocation-location solution. A V(s) value is computed and recorded for each solution. The one with the lowest V(s) value is chosen as the final solution. This swap step ends when all possible coupling of facility locations and facilities are exhausted.

3.2 Insufficient Capacity

The research also addresses insufficient capacitated problems, which were not considered in CAPCLUST. The procedure for insufficient capacity is similar to that for sufficient capacity but for two modifications.

First, the priority value for a demand point is the distance to its closest facility (instead of the difference in distances to its first and second closest unfulfilled facilities, as used in Section 3.1). This new calculation of priority values essentially assigns demand points to their nearest facility whenever possible. Second, it incorporates an extra step (*CapacityAnnealing*) after the allocation-location iteration with the initial location set (Step 2 – see previous section) and before swapping the facility locations (Step 3 – see previous section). Given the facility locations produced by *AllocationAndLocation*

(Step 2), facility capacities are decreased gradually and proportionally until the actual capacities are met during the *CapacityAnnealing* step. Each decrease in the capacity is followed by an iterative allocation-location to fine tune the facility locations. The *CapacityAnnealing* step stops once the actual capacities are reached. This additional step is devised to avoid serious deviation from optimal location results which can be caused by an abrupt and dramatic reduction in the capacities. The facility locations are then swapped and re-configured, same as the *FacilityLocationSwap* step (see Section 3.1) for sufficiently capacitated problems.

3.3 A GA-based Alternative to the Location Strategy

The difference between the clustering-based method and its GA alternative lies in the technique used to optimize facility locations. The clustering-based method optimized locations via the K-means technique while the GA method optimizes facility locations via GA operations. The GA alternative consists of two closely coupled components, i.e. a capacity-constrained allocation component and an adapted GA component. The capacity-constrained allocation component is the same as the allocation component in the clustering method. The GA component generates and optimizes facility locations by transforming the location subproblem to a simplified representation (e.g. a sequence of bits) and a simple set of operators (e.g. production, mutation, and crossover). For a specific facility location solution provided by the GA, the allocation component is used to assign demand points to the facilities under capacity constraints and calculate its V(s) measure. This measure is the fitness function used by the GA component to rank candidate solutions. Similar to the clustering-based method, the GA alternative is also capable of dealing with insufficiently capacitated and sufficiently capacitated problems. The overall procedure for the GA method is as follows:

- **(Step 1)** **FacilityInitialization**: Initialize facility locations by randomly choosing locations from the demand locations.
- **(Step 2)** **LocationEvaluation**: For the given facility locations: (a) use the allocation component to assign demand points to facilities under capacity constraints; (b) compute and record its V(s) value; (c) swap facility locations (needed when facility capacities are different). Repeat (a–c) until facilities are coupled with the given locations in all possible ways. A set of the lowest V(s) values and their corresponding solutions are recorded.
- **(Step 3)** **GenerationEvolution**: Using the above best-ranked candidates to breed a new generation of solutions using GA operations. Then repeat Step 2 until the maximum of generations is reached.
- **(Step 4)** **OptimalSolution**: Choose the solution with the lowest V(s) among all as the final solution.

Similar to the clustering-based method, the GA alternative starts with an initial location set (step 1). A collection (generation) of location sets is generated and evaluated by the constrained allocation component (step 2). The facility locations are then swapped exhaustively with fixed coupling of capacities and facilities. The allocation process follows each swap and V(s) is computed. The best location set is updated whenever a lower V(s) is observed. If the capacities are all the same, the swapping of facility locations is unnecessary and skipped. Next, new generations are produced via GA operations and evaluated (step 3). An existing open-source genetic algorithm package, JGAP (Meffert et al. 2006), is adapted to perform GA operations. The chromosome representation and the fitness function are customized to fit CFLP. Specifically, each chromosome consists of m distinct facility locations, which is initialized as a subset of the n demand points. We also used the V(s) as the fitness function for the GA to rank candidate solutions. Thus the solutions generated by the GA-based method are comparable to those by the clustering-based method. The *GenerationEvolution* step terminates when the maximum of generations is reached. The solution with the lowest V(s) value is the final solution.

4 Evaluation Experiments

The above clustering-based location-allocation method and its GA variant are tested with a variety of datasets, including a synthetic data set of random spatial distribution of demands, two synthetic data sets with spatially clustered demands, and a real data set. Facility capacities are different from each other in all datasets. Without losing generality, we assume that the demand at each point location in the synthetic data is the same (set as 1) to simplify the presentation of experiment results. In the real data set, the demand for each point location can be different. The experiments include scenarios of both sufficient total capacity and insufficient total capacity.

Table 2 presents a summary of the experiment results. The first three columns indicate the size of the dataset (# *demand points*/# *facility locations*), the spatial distribution type of demand locations (random, clustered, real), and the sufficiency of the overall capacity (sufficient or insufficient). The next two columns show the execution time (in seconds) and the V(s) value of the final solution by the clustering-based method. The last two columns are the execution time (in seconds) and the V(s) of the GA-based final solution with varying numbers of generations (i.e. 100, 300, and 500). All experiments are carried out on a desktop computer using the Microsoft Windows XP operation system, a Pentium 4 CPU (3.40GHz) and 2 GB of RAM memory. For the clustering method, the maximum number of allocation-location iterations allowed is 3,000, which is good enough for deriving stable location results with affordable computational cost. For the adapted GA alternative, each generation contains 50 location sets (chromosomes).

Table 2 Comparison on the synthetic datasets and the real dataset

Data set			clustering		GA (Generations = 100/300/500)	
Size (n/m)	Demand Distribution	Suf*	T	V(S)	T	V(S)
1500/5	Random	Y	12	0.1854	20/51/139	0.1891/0.1903/0.1915
1500/5	Clustered	Y	48	0.0850	18/42/60	0.0896/0.0853/0.0851
1500/5	Clustered	N	17	0.0774	16/36/59	0.0781/0.0806/0.0772
867/5	Real	N	272	0.3269	252/625/1050	0.3157/0.3126 /0.3120

* Indicating whether the overall capacity is adequate ("Y") or inadequate ("N")

4.1 Experiments with a Random Spatial Distribution

Figures 2a–c illustrate the searching process of the clustering approach for a data set of random spatial distribution and sufficient overall capacity. Each point location has a demand (population) of 1. The capacities of the five facilities (labeled "A", "B", "C", "D", and "E") are 100, 200, 300, 400 and 500, respectively. The capacity/demands served are shown after each facility label. Facilities that are filled up to their capacity are represented by filled rectangles and facilities that are not full are represented by filled circles. Demand locations assigned to the same facility are shown in the same color.

At the very beginning, the locations of the five facilities are randomly chosen (Figure 2a). With increased capacities, the iterative allocation-location procedures find a good set of facility locations (Figure 2b). Next, these locations are swapped among facilities

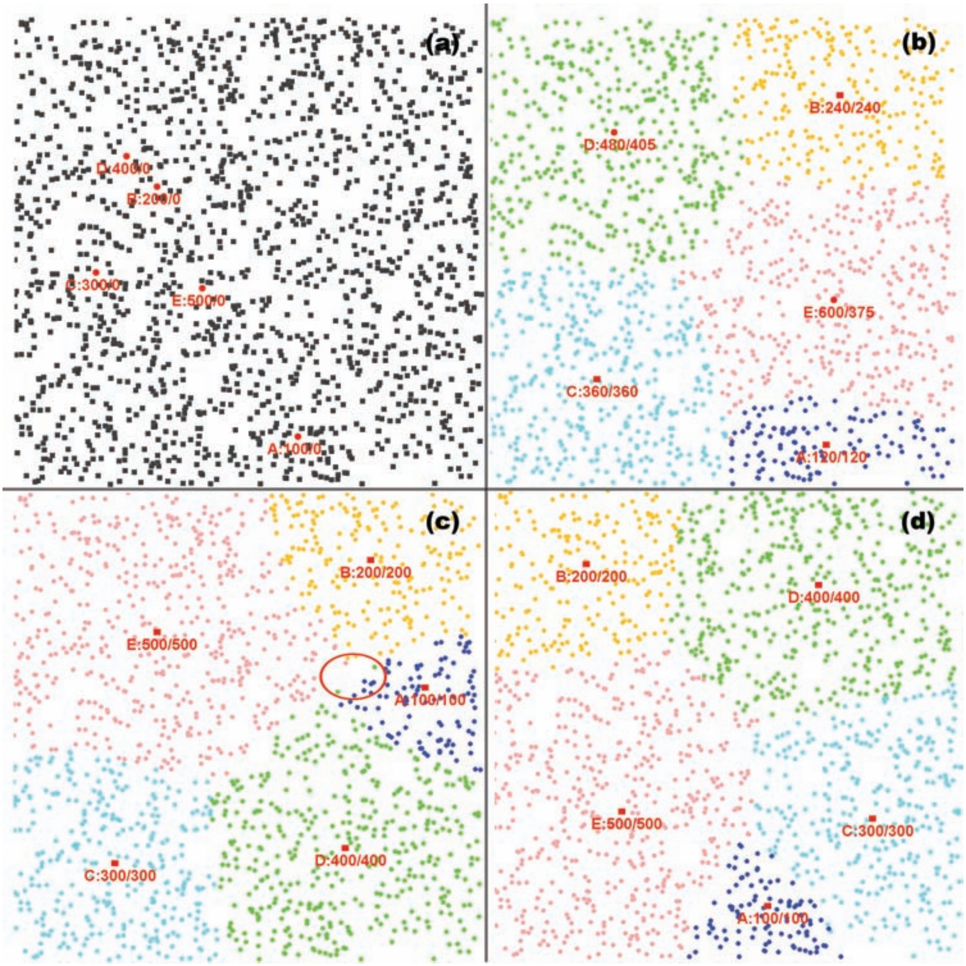

Figure 2 Solutions to a randomly distributed data set with sufficient overall capacity: (a) random initialization; (b) intermediate K-means solution; (c) final K-means solution; and (d) GA-100 solution

with the original capacity values, since the five capacities are different. The demand points are reassigned to facilities after each capacity swap. Figure 2c presents the final solution with the minimum V(s). A few point locations (in green and circled in Figure 2c) are not assigned to their closest facility ("A"). This can be caused by their relative low priority values. Intuitively, a better solution can be achieved by reassigning these points to facility A and reassigning the same number of locations located along the coverage boundary of A and D but associated with A to facility D. Figure 2d presents the GA solution derived with 100 generations, which does not place the facilities at the same locations as the clustering solution. The clustering-based method is more efficient in terms of the time cost and yet produces a better solution than the GA approach (Table 2). Note that the GA does not necessarily yield a better solution by evolving longer.

4.2 Experiments with a Clustered Spatial Distribution

Figure 3a shows the clustering solution and Figures 3b–d show the GA solutions with different numbers of generations (i.e. 100, 300, and 500 generations) for a clustered

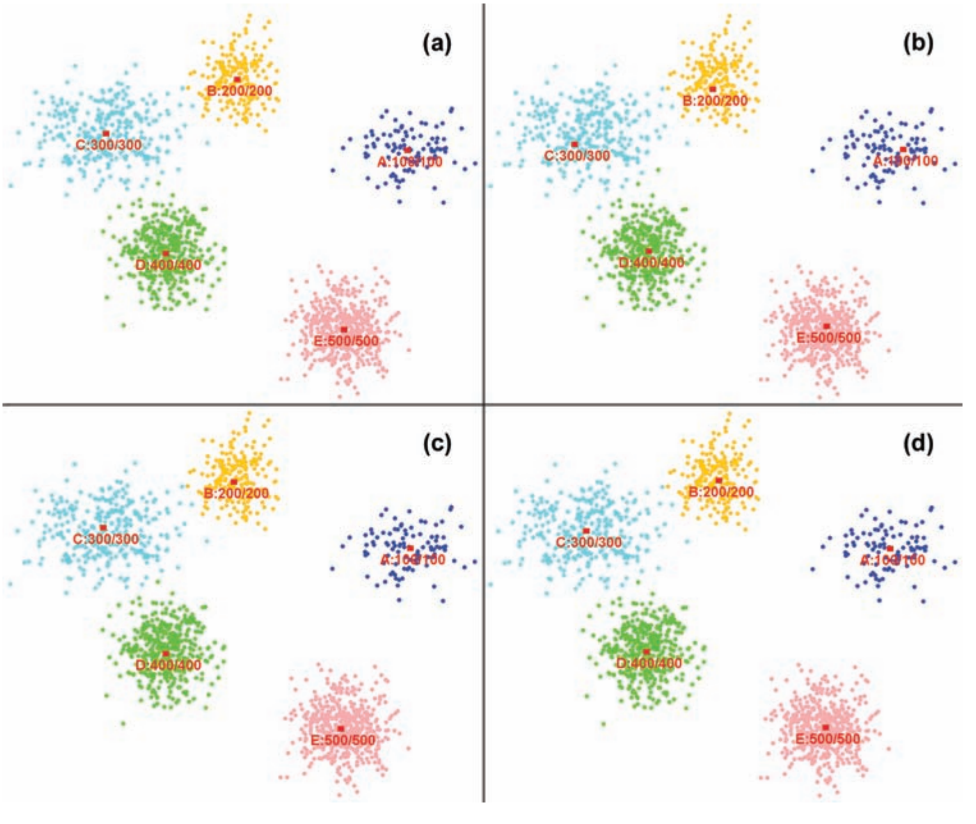

Figure 3 Solutions to a clustered distributed data set with sufficient overall capacity: (a) final K-means solution; (b) GA-100 solution; (c) GA-300 solution; and (d) GA-500 solution

synthetic data set. This synthetic dataset consists of five clusters of demand points, which contain 100, 200, 300, 400 and 500 points, respectively. The capacities of the five facilities to be located are also 100, 200, 300, 400, and 500. The demand of each point location is set as 1. The center and the diameter of each cluster are chosen randomly within a range and thus some clusters may overlap.

The clustering method locates each facility at the center of a demand cluster of the same size and assigns each demand location to its closest facility (Figure 3a). The GA with 100 (Figure 3b) generations positions each facility within a demand cluster of the same size as the clustering method does but fails to adjust to the clustered distribution of the demand locations completely. Apparently, facility B can be moved upward to the center of the demand cluster (in yellow) for a lower V(s). The GA with 300 generations (Figure 3c) also fails to locate all the facilities at the centers of demand clusters, though to a lesser degree. The GA with 500 generations (Figure 3d) reaches the solution that is nearly identical to the clustering-based method using more time (Table 2). Note GA operations are only used to configure the locations of the five facilities while the allocation component (as used in the clustering-based method) is used to assign demands to facilities (once their locations are determined by the GA).

4.3 Experiments with Insufficient Capacity

A clustered spatial distribution of demands is used to evaluate and compare the two methods in an insufficiently capacitated scenario. Figure 4 presents the solution generated by the clustering-based method (Figure 4a) and the best GA solution (with 500 generations). The data set consists of five clusters, which contain 100, 200, 300, 400 and 500 demand points, respectively. The demand at each point location is 1. The capacities of the facilities are 80, 160, 240, 320, and 400 (20% less than the cluster sizes). The clustering-based method first proportionally increases each capacity so that the total capacity is sufficient to meet all demands. During the capacity annealing process, the

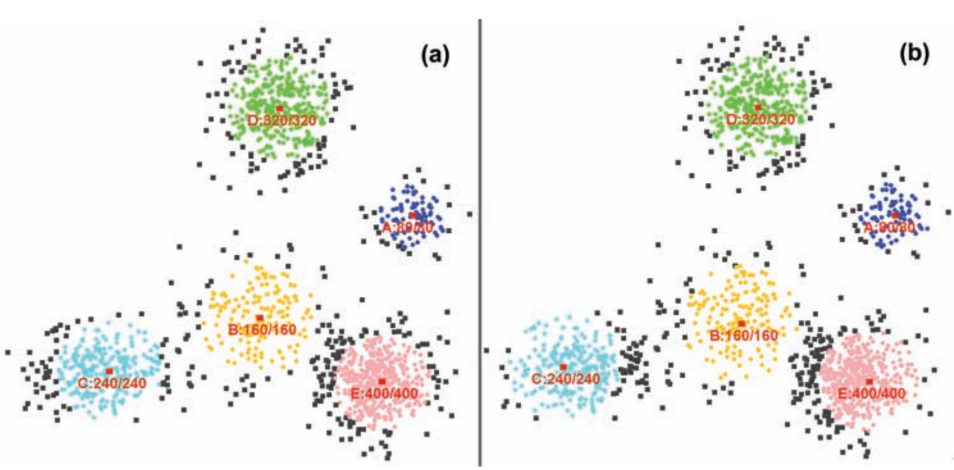

Figure 4 Solutions to a clustered distributed data set with insufficient overall capacity: (a) final K-means solution and (b) GA-500 solution

capacities are decreased gradually and the positions of the facilities are refined. When the result converges, separate and compact facility coverage is derived to cover the densest areas of demands (Figure 4a). The GA with 500 generations yields a similar allocation result as the clustering-based method using much longer time than the clustering method (Table 2). In another experiment conducted with a 10 times larger dataset (i.e. 15,000 demand points) than the above one, the clustering solution outperforms the GA solution with a lower V(s) value (0.078 vs. 0.080) and comparative time cost (21 minutes vs. 18 minutes).

4.4 Experiments with Real Data

This section presents an insufficient-capacity experiment with a real data set, which is to position five facilities and determine their service areas in South Carolina. The data set consists of 867 census tracts (Census 2000), with each treated as a demand location. The population of a census tract is considered as its demand, which varies from 197 to 16,745. The total population (4,212,012) is considered as the overall demand. Centroids of these census tracts are considered as the demand points, which are used to calculate the Euclidean distances between demand locations and facilities. Facility capacities range from 223,000 to 1,114,000. The overall capacity is about 20% less than the total demand.

Figure 5 presents the clustering-based solution (Figure 5a) and the GA solution after 500 generations (Figure 5b). The total population (in thousands) served by each facility is labeled. It is common to both solutions that: (1) coverage boundaries mainly cross rural areas; and (2) nearly all facilities are positioned close to most populated urban areas. This indicates both methods successfully adapt to the spatial distribution of demands. The best GA solution obtained after 500 generations (0.312, Table 2) is slightly better (4.6%) than the clustering-based solution (0.327). This can be caused by a dilemma of the clustering method during the searching process: it tries to force one facility (the one to the east, with a capacity of 446K) to cover two urban areas (Florence and Myrtle Beach). Therefore, the facility is located in the middle of them and thus affects the final quality measure. However, the clustering-based method is computationally much more efficient (272 seconds, Table 2) than the GA with 500 generations (1,050 seconds).

5 Discussion and Conclusions

The research presented here transforms a special location-allocation problem into a clustering problem. The proposed method is essentially a constrained K-means clustering method that indirectly optimizes the location-allocation quality under the individual and overall capacity restrictions. Since the allocation strategy adopted from the K-means algorithm can ensure a near-optimal allocation result (in terms of the objective function) when facility locations are fixed, this research focused more on designing methods to obtain a high-quality configuration of facility locations. This primary objective is accomplished through a suite of heuristics (i.e. capacity adjustment, capacity annealing, and facility-location swap) and the use of the sorted list based on priority value (during the allocation process). Experiments illustrate that theses strategies are successful in attaining good approximations of global optima of facility locations and an overall location-allocation solution of high quality.

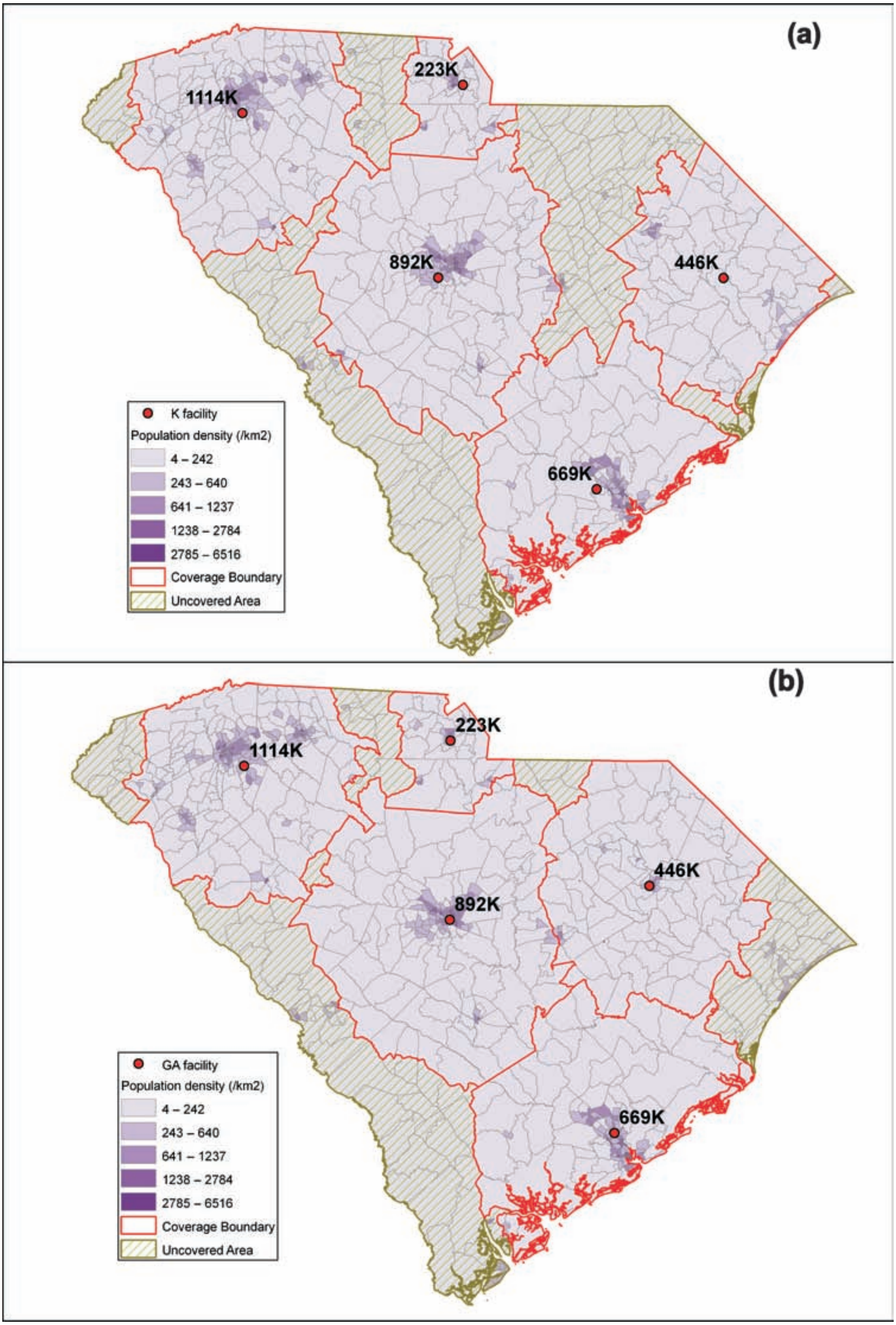

Figure 5 Solutions to a real data set with insufficient overall capacity: (a) final K-means solution and (b) GA-500 solution

Experiments are carried out to test the clustering-based approach against its variant, a combination of the constrained allocation component (in the clustering method) and a customized GA method (a global searching heuristic). Evaluation results show that the clustering-based approach is efficient (in terms of time cost) and produces comparable or even better solutions. As a deterministic approach, the clustering-based method involves fewer parameters than the GA-based method and is more stable.

There are several possibilities to improve the proposed clustering method. First, the swapping of facility locations can be improved. Because each swap is followed by allocation-location iteration and the number of swaps increases exponentially with the number of facilities, a more efficient strategy (e.g. based on heuristics) to swap the facility locations can save considerable computational cost. Second, modifications can be made to accommodate additional constraints or conditions other than the capacity concerns, such as the maximum distance that is allowed between the facility and the demand point, facility setup cost, and physical barriers. In these cases, the allocation component can be transformed into a flow network problem, which is more capable of handling complex constraints and objective functions. Third, visualization techniques should be integrated to allow users to interactively examine solutions, compare methods, and explore potential improvements.

Note

1. The first author was awarded the *Transaction in GIS* Student Paper Award for an earlier version of this manuscript that was submitted to and presented at the Summer Assembly of the University Consortium of Geographic Information Science held in Vancouver, Washington, 28 June to 1 July, 2006.

Acknowledgements

This research is partially funded by the United States Department of Homeland Security through the National Consortium for the Study of Terrorism and Responses to Terrorism (START), grant no. N00140510629. However, any opinions, findings, and conclusions or recommendations in this document are those of the authors and do not necessarily reflect the views of the U.S. Department of Homeland Security. The authors want to thank Dr. Frank Hardisty and two anonymous reviewers for insightful comments and useful suggestions.

References

Ahuja R K, Magnanti T L, and Orlin J B 1993 *Network Flows: Theory, Algorithms, and Applications*. Englewood Cliffs, NJ, Prentice Hall

Baldacci R, Hadjiconstantinou E, Maniezzo V, and Mingozzi A 2002 A new method for solving capacitated location problems based on a set partitioning approach. *Computers and Operations Research* 29: 365–86

Bertsimas D and Sim M 2003 Robust discrete optimization and network flows. *Mathematical Programming* 98: 49–71

Ceselli A and Righini G 2005 A branch-and-price algorithm for the capacitated p-median problem. *Networks* 45(3): 125–42

Chudak F A and Williamson D P 2005 Improved approximation algorithms for capacitated facility location problems. *Mathematical Programming* 102: 207–22

Densham P J and Rushton G 1992 Strategies for solving large location-allocation problems by heuristic methods. *Environment and Planning A* 21: 499–507

Díaz J A and Fernández E 2002 A branch-and-bound algorithm for the single source capacitated plant location problem. *European Journal of Operational Research Society* 53: 728–40

Díaz J A and Fernández E 2005 Hybrid scatter search and path relinking for the capacitated p-median problem. *European Journal of Operational Research* 169: 570–85

Duda R O, Hart P E, and Stork D G 2001 *Pattern Classification* (Second Edition). New York, John Wiley and Sons

Ehrgott M 2002 Location of rescue helicopters in South Tyrol. *International Journal of Industrial Engineering – Theory, Applications and Practice* 9: 16–22

Ernst A T and Krishnamoorthy M 1999 Solution algorithms for the capacitated single allocation hub location problem. *Annals of Operations Research* 86: 141–59

França P M, Sosa N M, and Pureza V 1999 An adaptive tabu search algorithm for the capacitated clustering problem. *International Transactions in Operational Research* 6: 665–78

Garey M R and Johnson D S 1979 *Computers and Intractability: A Guide to the Theory of NP-Completeness*. San Francisco, CA, W H Freeman

Goodchild M F 1984 ILACS: A location-allocation model for retail site selection. *Journal of Retailing* 60: 84–100

Guo D, Peuquet D J, and Gahegan M 2003 ICEAGE: Interactive clustering and exploration of large and high-dimensional geodata. *GeoInformatica* 7: 229–53

Haghifam M R and Shhabi M 2002 Optimal location and sizing of HV/MV substations in uncertainty-load environment using genetic algorithm. *Electric Power Systems Research* 63: 37–50

Han J, Kamber M, and Tung A K H 2001 Spatial clustering methods in data mining. In Miller H J and Han J (eds) *Geographic Data Mining and Knowledge Discovery*. London, Taylor and Francis: 188–217

Horner M W and O'Kelly W E 2005 A combined harvesting and transport planning within a sugar value chain. *Journal of the Operational Research Society* 57: 367–76

Jain A K and Dubes R C 1988 *Algorithms for Clustering Data*. Englewood Cliffs, NJ, Prentice Hall

Jain A K, Murty M N, and Flynn P J 1999 Data clustering: A review. *ACM Computing Surveys* 31: 264–323

Kalvenes J and Kennington E O 2005 Hierarchical cellular network design with channel allocation. *European Journal of Operational Research* 160: 3–18

Kohler J A 1973 TORNQVIST: Heuristic solution to the M-center location-allocation problem. In Rushton G, Goodchild M F, and Ostresh L M (eds) *Computer Programs for Location-allocation Problems*. Iowa City, IA, The University of Iowa, Department of Geography Monograph No 6

Kratica J, Stanimirovic Z, Tosic D, and Filipovic V 2005 Genetic algorithm for solving uncapacitated multiple allocation hub location problem. *Computing and Informatics* 24: 414–26

Lorena L A N and Senne E L F 2004 A column generation approach to capacitated p-median problems. *Computers & Operations Research* 31: 863–76

Maniezzo V, Mingozzi A, and Baldacci R 1998 A bionomic approach to the capacitated p-median problem. *Journal of Heuristics* 4: 263–80

Marín A and Pelegrín B 1997 A branch-and-bound algorithm for the transportation problem with location of p transshipment points. *Computers and Operations Research* 24: 659–78

Meffert K, Meskauskas A, Meseguer J, and Martí E D 2006 Java Genetic Algorithms Package (JGAP, Ver. 3.0, RC2). WWW document, http://jgap.sourceforge.net

Mulvey J M and Beck M P 1984 Solving capacitated clustering problems. *European Journal of Operational Research* 18: 339–48

Murray A T and Gerrard R A 1997 Capacitated service and regional constraints in location-allocation modeling. *Location Science* 5: 103–18

Nauss R M 1978 An improved algorithm for the capacitated facility location problem. *The Journal of the Operational Research Society* 29: 1195–1201

Negreiros M and Palhano A 2006 The capacitated centred clustering problem. *Computers and Operations Research* 33: 1639–63

ReVelle C S and Swain R W 1970 Central facilities location. *Geographical Analysis* 2: 30–42

Scheuerer S and Wendolsky R 2006 A scatter search heuristic for the capacitated clustering problem. *European Journal of Operational Research* 169: 533–47

Sedeño-Noda A and González-Martín C 2001 An algorithm for the biobjective integer minimum cost flow problem. *Computers and Operations Research* 28: 139–56

Vinod H D 1969 Integer programming and the theory of grouping. *Journal of the American Statistical Association* 64: 506–19

Wu L-Y, Zhang X-S, and Zhang J-L 2006 Capacitated facility location problem with general setup cost. *Computers and Operations Research* 33: 1226–41