**Noname manuscript No.**
(will be inserted by the editor)

# Solution of Asymmetric Discrete Competitive Facility Location Problems using Ranking of Candidate Locations

**Algirdas Lančinskas · Julius Žilinskas · Pascual Fernández · Blas Pelegrín**

Received: date / Accepted: date

**Abstract** We address a discrete competitive facility location problem with an asymmetric objective function and a binary customer choice rule. Both an integer linear programming formulation and a heuristic optimization algorithm based on ranking of candidate locations are designed to solve the problem. The proposed population-based heuristic algorithm is specially adapted for the discrete facility location problems by using their features such as geographical distances and the maximal possible utility of candidate locations, which can be evaluated in advance. Performance of the proposed algorithm was experimentally investigated by solving different instances of the model with real data of municipalities in Spain.

**Keywords** Asymmetric Facility Location · Binary Choice Rule · Combinatorial Optimization · Random Search · Population-based Heuristic Algorithms

## 1 Introduction

Facility location deals with finding in some sense the best locations for facilities providing goods or services in a geographical area. The right location for a facility depends on different factors such us type of service or product that the firm provides, supply chain characteristics, market environment, etc. A facility location problem usually is formulated as a mathematical optimization problems with an objective function which is subject to minimize or maximize; e.g. utility of the new facilities which is subject to maximize or any unwanted effect caused by locating the new facilities which is subject to minimize, etc.

Competitive facility location problems arise when firms provide goods or services to customers in a certain geographical area and compete for the market share with other firms. The location of their facilities is a key point to determine market share. There are various facility location models and strategies to solve such problems (see for instance [5, 10, 19, 21]), which differ by their ingredients such as a facility attraction function, customers behavior rules, decision variables, a search space, objective function(s), etc.

Many models in the literature deal with facility location for one firm, which enters the market by locating some new facilities that will compete with other firms already established in the market. The entering firm faces a mathematical optimization problem aimed at finding the optimal locations for the new facilities subject to maximization of their market share. Depending on whether the new facilities can be located anywhere in a given region or their locations must be selected from a finite set of candidate locations the firm deals with the continuous or discrete competitive facility location problem, respectively.

A customer behavior is defined by the rule that customers follow when they are choosing the facility to buy a service. In the literature, it is usual to consider the distance to the facility as the first criterion used by the customer, so that the demand is served by the closest facility [3, 23]. Some variants to model attractiveness of a facility are proposed by Huff [14], where attractiveness depends on the distance to the facility and its quality,

A. Lančinskas (*corresponding*), J. Žilinskas
Institute of Data Science and Digital Technologies,
Vilnius University, Akademijos 4, 08663, Vilnius, Lithuania
E-mail: algirdas.lancinskas@mii.vu.lt, julius.zilinskas@mii.vu.lt

P. Fernández, B. Pelegrín
Department of Statistics and Operations Research, University of Murcia,
Campus Espinardo, 30071 Murcia, Spain.
E-mail: pfdez@um.es, pelegrin@um.es

which can be estimated by the size of the facility, the number of parking spaces, leisure and/or entertainment areas, etc.

The most common customers behavior rules are called proportional and binary. If the behavior is based on the proportional rule, then the buying power of a customer is split among all facilities in proportion to their attraction [4,17,22]. In the case of the binary rule, the customer patronizes a single facility – the most attractive one [11,24,25].

A lot of facility location problems use a symmetric objective function – the value of which is equal for any permutation of values of variables. However, real-world applications usually correspond to asymmetric facility location problems, where the position of a facility in the solution is crucial; e.g. facility location problems, where the new facilities are associated with some different properties, such as their quality. Consider a firm (supermarkets, shopping centers, petrol stations, etc.) which can open three different types of facilities depending on their characteristics (size, parking, services, etc.), i.e., facilities with three possible quialities. If the firm decides to locate one facility of each type, since the attraction between customers and facilities directly depending on facility quality, the captured market share by the firm locating facilities of qualities $q_1$, $q_2$, and $q_3$ at points 1, 2, and 3, respectively, will be different of the captured market share if facilities of qualities $q_1$, $q_2$, and $q_3$ are located at points 2, 3, and 1, respectively. So, in this way, we consider that the objective function is an asymmetric function.

Solving a real-world CFLP is usually complex and computationally expensive due to reasons such as complex objective function(s) or need of complex analysis of a large amount of data, e.g. population of prospective customers, their current and expected behavior when choosing the facility for a service, etc. Due to these and similar reasons it can be impossible to find the optimal solution(s) within a reasonable time. Therefore, heuristic methods, which can be applied to approximate the optimal solution(s) of a specific optimization problem, are often used to tackle a real-world CFLP. See [1,2,18] for examples of application of heuristic algorithms for single and multi-objective facility location problems.

In this paper we will focus on the discrete facility location problem that uses the binary rule for the customer behavior. The goal is to choose the optimal locations for a set of new facilities from a finite set of candidate locations taking into account that the new facilities have different qualities which are fixed and known, what makes the problem to be asymmetric.

This work is continuation of our previous work focused on solution of symmetric facility location problems using a heuristic algorithm based on ranking of candidate locations. The algorithm is described in [7], where the idea of candidate location ranking was proposed. The ranking strategy was especially adopted to the discrete facility location problems by including some features of the problem such us geographical distances, what enables a kind of local search. Although the algorithm demonstrates good performance when solving various instances of facility location problems, the ranking strategy makes the algorithm limited to the symmetric facility location problems.

This paper is focused on extension of the strategy for ranking of the candidate locations which makes it suitable to deal with asymmetric facility location problems. The proposed strategy is used to design a new heuristic algorithm which, in contrast with the one proposed in [7], is population-based and uses a new strategy to reject without evaluation obviously non-acceptable candidate solutions.

The reminder of the paper is organized as follows. In Sect. 2, the notation and formulation of the discrete facility location problem is presented. The proposed algorithm to solve the problem is described in Sect. 3 and the results of experimental investigation of the algorithm are presented in Sect. 4. Finally, some conclusions are formulated in Sect. 5.

## 2 Facility Location Problem

An entering firm wants to locate some new facilities with pre-fixed qualities in a geographical region where similar facilities of other competing firms are already present, but for simplicity, we consider all competing firms as only one, the competitor. The entering firm wants to locate $s$ new facilities with given qualities $q_1, q_2, \ldots, q_s$. A solution to the problem will consist on a $s$-vector where its $i$-th component indicates where to locate a new facility with quality $q_k$, $k = 1, \ldots, s$. In traditional competitive location models, when new facilities are to be opened and the quality is not considered as a variable, or quality is supposed to be associated with location points, an exchange in the components of the solution vector leads to the same solution of the model. However, in this new model, where each position in the solution vector is associated with the facility quality, an exchange among the facility locations leads to a different solution. When this occurs, we are considering an asymmetric competitive facility location problem.

Customers are considered to be aggregated to geographic demand points which are spatially separated in order to make the problem computationally tractable (see [9]). It is assumed that customers' demand, qualities of the competitors' facilities and qualities of the new facilities are fixed and known.

To propose a formulation for this model, the following general notation is used:

Indices:

| | |
|---|---|
| $i, I$ | Index and set of demand points (customers) |
| $j, L$ | Index and set of candidate locations (discrete set) |
| $k, K$ | Index and set of indices of qualities |

Data:

| | |
|---|---|
| $w_i$ | Demand at demand point $i$. |
| $d_{ij}$ | Distance between demand point $i$ and facility $j$. |
| $a_{ijk}$ | Attraction that demand point $i$ feels for a facility located at $j$ with quality $q_k$. Let be $a_{ijk} = \frac{q_k}{1+d_{ij}}$. |
| $a_i(F)$ | Maximum attraction that $i$ feels for facilities in $F$, where each facility in $F$ is noted by its location point and quality. |
| $s$ | Number of new facilities to be located. |
| $Q^{(r)}$ | Set of $r$ quality values, $Q^{(r)} = \{q_1, q_2, \dots, q_r\}$. |
| $C$ | Set of existing facilities of competing firms. |

Variables:

| | |
|---|---|
| $X$ | Set of locations for the new facilities and its qualities, $X = \{x_{jk} : j \in L, k \in K\}$. |

We consider that customers follow a binary behaviour to choose the facility which serves its demand. So, the full demand of each customer will be satisfied by the facility with maximum attraction. It may occur that there is more than one facility with maximum attraction owned by the entering firm or the competitors. If all tied facilities are owned by the entering firm, the firm captures the full demand of the customer, while if none facility with maximum attraction owns to the entering firm, no demand is captured from the customer. Otherwise, the entering firm captures a fixed proportion of customer's demand.

This problem can be formulated as an ILP problem considering the following sets:

$$L_i^{>} = \{(j,k) \in L \times K : a_{ijk} > a_i(C)\}, \forall i \in I \tag{1}$$

$$L_i^{=} = \{(j,k) \in L \times K : a_{ij} = a_i(C)\}, \forall i \in I \tag{2}$$

$$I^* = \{i \in I : L_i^{>} \cup L_i^{=} \neq \emptyset\}, \forall i \in I \tag{3}$$

Since the entering firm only will capture demand of a customer $i$ if its attraction for the new facilities is at least equal to the attraction for competitor's facilities, the set $I^*$ includes the customers which demand can be total or partially captured by the entering firm. To introduce a formulation of the model as a binary programming problem, the following variable sets are also considered:

$$x_{jk} = \begin{cases} 1, \text{ if a new facility is located at } j \\ \quad \text{with quality } q_k; \\ 0, \text{ otherwise, where} (j,k) \in L \times K \end{cases}$$

$$y_i = \begin{cases} 1, \text{ if customer } i \text{ is fully captured} \\ \quad \text{by the entering firm;} \\ 0, \text{ otherwise, where } i \in I^* \end{cases}$$

$$z_i = \begin{cases} 1, \text{ if customer } i \text{ is partially captured} \\ \quad \text{by the entering firm} \\ 0, \text{ otherwise, where } i \in I^* \end{cases}$$

Then, the asymmetric competitive location problem can be formulated as:

$$(P) \begin{cases} \max & \sum_{i \in I^*} w_i y_i + \sum_{i \in I^*} \theta_i w_i z_i \\ \text{s.t.} & y_i + z_i \leq 1, i \in I^* & (4) \\ & y_i \leq \sum_{(j,k) \in L_i^{>}} x_{jk}, \forall i \in I^* & (5) \\ & z_i \leq \sum_{(j,k) \in L_i^{=}} x_{jk}, \forall i \in I^* & (6) \\ & \sum_{(j,k) \in L \times K} x_{jk} = s & (7) \\ & \sum_{k \in K} x_{jk} \leq 1, \forall j \in L & (8) \\ & \sum_{j \in L} x_{jk} = 1, \forall k \in K & (9) \\ & x_{jk} \in \{0,1\}, \forall (j,k) \in L \times K, \\ & y_i, z_i \in \{0,1\}, \forall i \in I^* \end{cases}$$

where $\theta_i$ is the proportion of demand captured by the entering firm from customer $i$ in case of ties on maximum attraction between the competitor and the new facilities; $\theta_i \in [0,1]$, being $\theta_i = 0$ if customers in $i$ are conservative, and $\theta_i = 1$ if are novelty oriented. The usual value is $\theta_i = 0.5 \forall i$, which means that the entering firm captures a half of customer demand in case of tie. The objective function gives the total demand captured by the entering firm, which first term is the demand due to fully captured customers and the second one is due to partially captured customers. Constraints set (4) ensures that each customer can only be totally or partially captured. Constraints set (5) imposes that a customer is fully captured only if there exists some new facility located at $j$ with quality $q_k$ so that $(j,k) \in L_i^{>}$ (analogously for constraints set (6) with partially captured and $(j,k) \in L_i^{=}$). Constraint (7) limits the number of new facilities to $s$. Constraints set (8) stays that only a new facility can be located at each candidate location $j$, and constraints set (9) ensures that a new facility will be located for each given quality. The previous formulation extends the classical formulation with one index variables of the symmetric competitive location model with binary customer choice rule. In the new model not only the locations for the new facilities must be decided, but also the quality of the facility at each location. The previous formulation can be simplified as follows.

Note that if two or more new facilities were located at the same location candidate point, as customers follow a binary choice rule, only the facility with the biggest quality would be chosen to serve its demand, and if all these facilities had the same quality, the entering firm would capture the same demand locating only one of them. As the objective is to maximize the total captured demand, an optimal solution would never contain more than one new facility at the same point. This means that constraints set (8) is unnecessary and can be removed. On the other hand, constraints set (9) can be rewritten in a less restrictive way as a less or equal inequality, because at any optimal solution of the problem, this constraint will be verified in equality for each one of the prefixed qualities, otherwise this would mean that any of the new facilities is not located any more, which would imply that constraint (7) does not hold. Furthermore, each constraint $y_i, z_i \in \{0,1\}$, $\forall i \in I^*$, can be replaced by $y_i, z_i \geq 0$, $\forall i \in I^*$. If so, in the optimal solution $y_i, z_i$, $\forall i \in I^*$, will take the value 0, or the maximum possible value which is 1 due to constraints set (4). Then, the formulation of $(P)$ is equivalent to:

$$\begin{cases} \max \displaystyle\sum_{i\in I^*} w_i y_i + \sum_{i\in I^*} \theta_i w_i z_i \\ \text{s.t.} \;\; y_i + z_i \leq 1, i \in I^* \\ \qquad y_i \leq \displaystyle\sum_{(j,k)\in L_i^{>}} x_{jk}, \forall i \in I^* \\ \qquad z_i \leq \displaystyle\sum_{(j,k)\in L_i^{=}} x_{jk}, \forall i \in I^* \\ \qquad \displaystyle\sum_{(j,k)\in L\times K} x_{jk} = s \\ \qquad \displaystyle\sum_{j\in L} x_{jk} \leq 1, \forall k \in K \\ \qquad x_{jk} \in \{0,1\}, \forall (j,k) \in L\times K, y_i, z_i \geq 0, \forall i \in I^* \end{cases}$$

## 3 Ranking-based Discrete Optimization Algorithm

The concept of Ranking Based Discrete Optimization Algorithm (RDOA) has been proposed in [7]. The initial version of the algorithm was based on a single-agent random search strategy, where new solutions were always generated by modifying the best known solution. The new version of RDOA, being presented in this section, is population-based where new solutions are generated by modifying a solution

$$X = \{x_1, x_2, \dots, x_s\} \tag{7}$$

sampled from a set $P$ of the best solutions found so far. At the beginning of the algorithm the set $P$ contains a single randomly generated solution, but is subject to be supplemented up to $n_P$ elements.

When $P$ contains more than one item, then $X$ is sampled considering sampling probability proportional to its objective function value; the better solution – the larger probability to be sampled.

A new candidate solution

$$X^{(n)} = \{x_1^{(n)}, x_2^{(n)}, \dots, x_s^{(n)}\} \tag{8}$$

is generated by in turn taking candidate locations from the best known solution $X$ and changing them to another ones randomly selected from the set of all possible candidate locations excluding those which already forms $X$ or $X^{(n)}$. Each location $x_i$ is changed with probability $1/s$, and is copied without changing with probability $1-1/s$:

$$x_i^{(n)} = \begin{cases} l \in L \setminus (X \cup X^{(n)}), & \text{if } \xi_i < 1/s, \\ x_i, & \text{otherwise}, \end{cases} \tag{9}$$

where $\xi_i$ is a random number uniformly generated over the interval $[0,1]$, and $i = 1, 2, \dots, s$.

A candidate location $l_j \in L$ can be selected to represent $i$-th location in $X^{(n)}$ with probability

$$\pi_{ij}^{(r)} = \frac{r_{ij}}{\sum_{k=1}^{|L|} r_{ik}}, \tag{10}$$

where $r_{ij}$ is a rank of candidate location $l_j$ being as a place for $i$-th new facility. Analogously, sampling probability of $l_j$ as a place for $i$-th new facility can be evaluated by including geographical distance between candidate location $l_j$ and current location of $i$-th new facility:

$$\pi_{ij}^{(rd)} = \frac{r_{ij}}{d(l_j, x_i) \sum_{k=1}^{|L|} \frac{r_{ik}}{d(l_k, x_i)}}, \tag{11}$$

where $d(l_j, x_i)$ is a geographical distance between candidate location $l_j \in L$ and candidate location $x_i \in X$ which is being changed $(i = 1, 2, \dots, s)$.

Lets denote by

$$R_i = (r_{i1}, r_{i2}, \ldots, r_{ij}, \ldots) \tag{12}$$

the ranks of all candidate locations from $L$ being as a place for $i$-th new facility. Then all ranks can be described by matrix $R$ of $s$ rows and $|L|$ columns:

$$R = \begin{pmatrix} R_1 \\ R_2 \\ \vdots \\ R_s \end{pmatrix} = \begin{pmatrix} r_{11} & r_{12} & \ldots & r_{1|L|} \\ r_{21} & r_{22} & \ldots & r_{2|L|} \\ \vdots & \vdots & \ddots & \vdots \\ r_{s1} & r_{s2} & \ldots & r_{s|L|} \end{pmatrix}. \tag{13}$$

Initial values of $R$ are equal to 1 and is dynamically adjusted with respect to success and failures when generating a new candidate solution. If the newly generated solution $X^{(n)}$ is better than the worst solution $X^{(w)} \in P$, then

(1) $P$ is updated by including $X^{(n)}$:

$$P \leftarrow P \cup \{X^{(n)}\}. \tag{14}$$

If size of $P$ exceeds its size limit $n_P$, then the worst candidate solution $X^{(w)}$ is removed from $P$.

(2) ranks of all candidate locations are increased by one:

$$r_{ij} = \begin{cases} r_{ij} + 1, & \text{if } l_j = X_i^{(n)}, \\ r_{ij}, & \text{otherwise,} \end{cases} \tag{15}$$

(3) corresponding ranks of all candidate locations which form a candidate solution in $P$ which is improved by $X^{(n)}$, but do not form it are reduced:

$$r_{ij} = r_{ij} - k, \tag{16}$$

where

$$k = |\{X \subset P : x_i = l_j \wedge M(X) < M(X^{(n)})\}|. \tag{17}$$

If $X^{(n)}$ do not improve the worst solution $X^{(w)} \in P$, then the ranks of all candidate locations forming unsuccessfully generated solution $X^{(n)}$, but do not form the worst candidate solution $P$, are reduced by one:

$$r_{ij} = \begin{cases} r_{ij} - 1, & \text{if } l_j = x_i \wedge l_j \neq x_i^{(w)} \\ r_{ij}, & \text{otherwise.} \end{cases} \tag{18}$$

Reducing rank values can make a rank equal to zero or negative value, e.g. $r_{ij} = -k$, where $k \geq 0$. Then all ranks in $R_i$ are increased by $k + 1$.

After processing the newly generated solution algorithm continues to the next iteration, where another solution $X$ is sampled from $P$ to generate a new solution $X^{(n)}$. The procedure continues till stopping criterion is satisfied, which is based on the number of function evaluation.

The initial pool size $n_P$ is given as an algorithm parameter and is further reduced thus letting to perform a wider exploration of the search space in the early stage of the algorithm and focus on local search in the final stage of the algorithm. The reduction is performed after every 20% of function evaluations by removing a half worst solutions.

The algorithm which uses the ranks only to evaluate sampling probability for candidate locations is denoted by RDOA and the algorithm, which additionally includes geographical distance is denoted by RDOA-D.

### 3.1 Pre-calculated Market Share

Let's denote by $m_{ij}^{(1)}$ the market share of $i$-th new facility located in candidate location $l_j$ when locating a single facility, and by $m_{ij}^{(s)}$ – the market share of $i$-th new facility located in $l_j$ when locating $s > 1$ new facilities.

Evaluation of $m_{ij}^{(1)}$ requires approximately $s$ times less computational effort than evaluation of $m_{ij}^{(s)}$. Although computation of $m_{ij}^{(1)}$, when $i = 1, 2, \ldots s$ and $j = 1, 2, \ldots |L|$, requires to devote a significant number of function evaluations, but the obtained information can be useful to save computational effort in later stages of the algorithm.

The first feature derived from $m_{ij}^{(1)}$ is market share obtained by a single facility located at $j$-th candidate location. Naturally, facilities that separately attract more customers should form better solution when locating $s > 1$ new facilities. Therefore $m_{ij}^{(1)}$ can be included when calculating sampling probability for candidate location $l_j$ in (9). Then sampling probability, expressed by (11) can be evaluated by

$$\pi_{ij}^{(rdp)} = \frac{r_{ij} m_{ij}^{(1)}}{d(l_j, x_i) \sum_{k=1}^{|L|} \frac{r_{ik} m_{ik}^{(1)}}{d(l_k, x_i)}}. \tag{19}$$

The market share $m_{ij}^{(1)}$ is evaluated taking into account the competition with preexisting facilities belonging to the competitors, but do not include competition between new facilities being located at the same time. Therefore,

$$m_{ij}^{(1)} \geq m_{ij}^{(s)}, \tag{20}$$

and

$$\sum_{l_j \in X} m_{ij}^{(1)} \geq M(X). \tag{21}$$

This information is useful when deciding whether a newly generated solution $X^{(n)}$ could improve the worst solution in the pool $P$ or not. If

$$\sum_{l_j \in X^{(n)}} m_{ij}^{(1)} \leq M(X^{(w)}), \tag{22}$$

then, due to (21), it is not possible that $X^{(n)}$ is attracts more customers than $X^{(w)}$ and there is no need to perform complete evaluation of the solution, which requires incomparably more computational effort than calculating sum of several numbers. This lets us save function evaluations for the donation of insignificant computational effort.

The algorithm, which, in addition to features of RDOA-D, includes pre-calculated market share in sampling probability evaluation and reject obviously unacceptable solutions, is denoted by RDOA-D-PreMS.

## 4 Experimental Investigation

The proposed algorithms have been experimentally investigated by solving asymmetric DCFLP aimed at location of new facilities of different qualities.

The database of 1000 municipalities in Spain were used as demand points. Ten preexisting facilities were located in 10 largest demand points with randomly generated qualities: the lowest quality – 45, the largest – 68, and the average – 56.

Instances of DCFLP vary on the number of candidate locations $|L| \in \{500, 1000\}$, and the number of new facilities $s \in \{3, 5\}$.

Different quality values for the new facilities were used. At the beginning it was expected to locate low quality facilities, with the corresponding quality values

$$Q_L^{(3)} = (10, 20, 30), \tag{23}$$

when $s = 3$ and

$$Q_L^{(5)} = (10, 20, 30, 40, 50), \tag{24}$$

when $s = 5$. Later the qualities were increased and

$$Q_H^{(3)} = (50, 60, 70) \tag{25}$$

were used to investigate the impact of the quality values.

Optimal solutions for different problem instances were found using deterministic optimization algorithm Xpress [8]. See Table 1, which will be used to check if an algorithm is able to find them or not, and if so, how many function evaluations are needed.

**Table 1** Optimal solutions found by deterministic algorithm Xpress.

| $\|L\|$ | $s$ | $Q$ | $X$ | $M(X)$ |
|---|---|---|---|---|
| 500 | 3 | $Q_L^{(3)}$ | (460, 41, 500) | 13.05 |
| 1000 | 3 | $Q_L^{(3)}$ | (622, 41, 500) | 13.05 |
| 500 | 3 | $Q_H^{(3)}$ | (4, 1, 2) | 51.45 |
| 1000 | 3 | $Q_H^{(3)}$ | (4, 1, 2) | 51.45 |
| 500 | 5 | $Q_L^{(5)}$ | (460, 81, 29, 331, 4) | 24.20 |

### 4.1 Average Market Share Obtained by Heuristic Algorithms

Independent on the heuristic algorithm, 5000 function evaluation was devoted for a single run. Due to stochastic nature of algorithms under investigation, 100 runs were performed for each experiment and average results were recorded. The maximal pool size was set to $n_P = 64$ and was reduced by a half after every 1000 function evaluations, thus using pool of 8 best candidate solutions for the last thousand of function evaluations.

The performance of the proposed algorithms were compared with the performance of Genetic Algorithm (GA), which was proposed by Holland in 1975 [13] and was successfully applied to combinatorial optimization problems (see [16,26]), including facility location problems (see [6,15]). GA is population-based algorithm, which simulates genetic such as crossover of individuals and mutation of derived individuals (see [20] for details).

The population size was set to 64 individuals – the same as initial pool size in the proposed algorithms. The uniform crossover with the rate equal to 0.8 and mutation rate equal to $1/s$ were used to generate new individuals.

The first experiment was focused on selecting optimal locations for 3 new facilities from a set of 500 and 1000 candidate locations considering low qualities $Q_L^{(3)}$ for the new facilities. All three versions of RDOA were used to approximate the optimal solution within 5000 function evaluations recording average market share of the best solution found after every 1000 function evaluations.

The results are presented in Figs. 1 and 2, where the the first figure presents results of the instance with 500 candidate locations and the second one – with 1000 candidate locations. The horizontal axis of a graph stands for the number of function evaluations and the vertical one – for the percent of market share captured by the new facilities.

One can see from the figures, RDOA without geographical distance and pre-calculated market share outperforms GA in both cases. Significant difference ap-

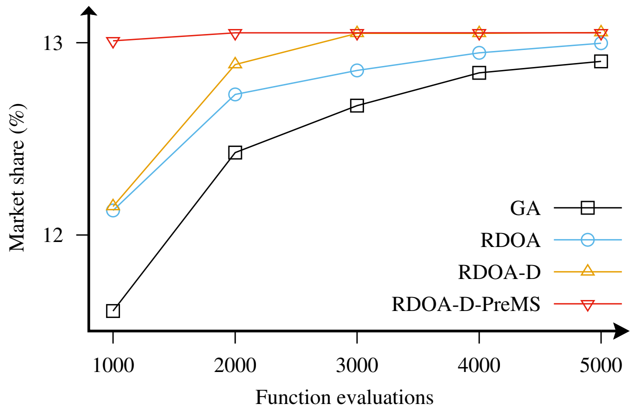


**Fig. 1** Results obtained when choosing locations for 3 new facilities from 500 candidate locations considering low qualities $Q_L^{(3)}$ for the new facilities.

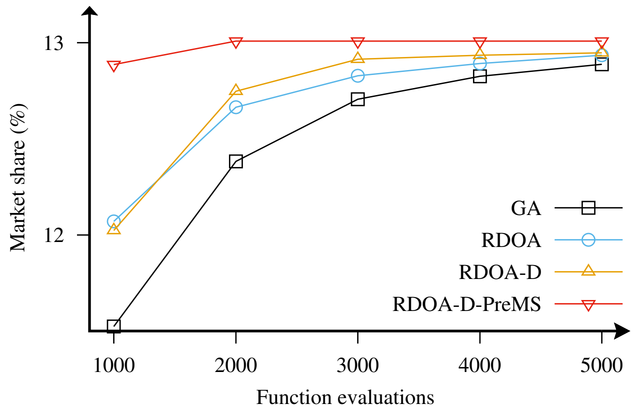


**Fig. 2** Results obtained when choosing locations for 3 new facilities from 1000 candidate locations considering low qualities $Q_L^{(3)}$ for the new facilities.

pears in early stage of the algorithm, what means that RDOA is able to find much more better solution in the beginning of the procedure, e.g. after 1000 function evaluations.

The advantage of inclusion of geographical distance in calculation of sampling probabilities for candidate locations (RDOA-D) is notable in later stage of the algorithm – it is specially notable in the instance with 500 candidate locations, where the best performance is achieved after 3000 function evaluations. Inclusion of the geographical distance is a kind of a local search and is more useful when updating a good candidate solution. This could be a reason for lower performance at the beginning of the algorithm.

The best performance was demonstrated by RDOA-D-PreMS, where both geographical distance and information obtained from pre-calculated market share were included. The average solution obtained by RDOA-D-PreMS after 1000 function evaluations is more less the same as average solution obtained by RDOA after 5000 function evaluations and notably better than average solution obtained by GA after 5000 function evaluations.

Duration of each experiment with RDOA-D-PreMS lasts from 5 to 8 seconds depending on a problem instance. The algorithm has additional computational work to calculate initial market-share of candidate locations, but the budget of function evaluations for further computations is reduced accordingly to keep the same total complexity of the algorithm. Some of solutions generated by RDOA-D-PreMS were rejected without function evaluation, therefore more solutions have to be generated in this algorithm. On the other hand, the main computational effort is devoted for function evaluation and generation of these unacceptable solutions takes insignificant amount of computational resources, which increases the duration of the algorithm by less than 3% of the total computational time.

## 4.2 Impact of the Problem Instance

Next experiment was aimed at investigation of impact of quality values for the new facilities to the performance of the optimization procedure. The qualities $Q_H^{(3)}$ were used as quality values for the new facility thus making average quality of a new facilities larger than average quality value of a preexisting facility.

The experiment was performed under the same conditions as the first one, but performance of the best version of the proposed algorithms – the RDOA-D-PreMS – was compared with GA in this experiment.

One can see from the results, presented in Figs. 3 and 4, the RDOA-D-PreMS significantly outperforms GA independent on function evaluations performed and instance of DCFLP. In both cases RDO-D-PreMS produces the best result after 2000 function evaluations, whereas the instance with 1000 candidate locations appeared to be more complicated for GA.

The third experiment was aimed at investigation of impact of the number of the new facilities to the performance of the algorithms. It was expected to choose locations for 5 new facilities from the set of 500 candidate locations considering low qualities $Q_L^{(5)}$ for the new facilities.

The results, presented in Fig. 5, demonstrate advantages of the proposed RDOA-D-PreMS – the average market share of the best solution found by RDOA-D-PreMS after 2000 function evaluations is quite close to the best market share of the best solution ever found

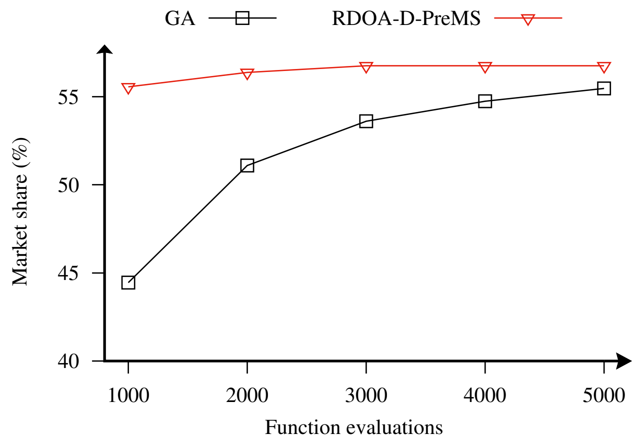


**Fig. 3** Results obtained when choosing locations for 3 new facilities from 500 candidate locations considering high qualities $Q_H^{(3)}$ for the new facilities.

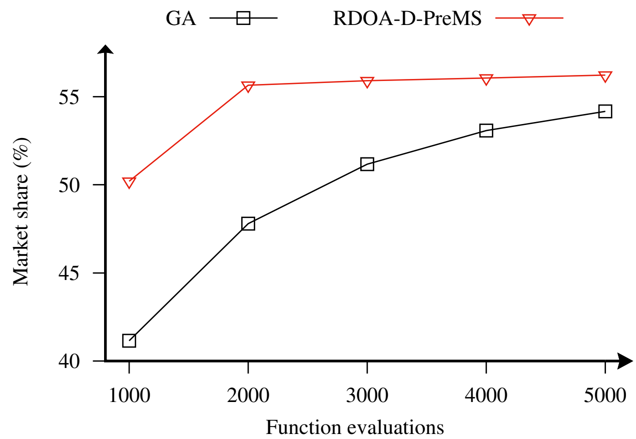


**Fig. 4** Results obtained when choosing locations for 3 new facilities from 1000 candidate locations considering high qualities $Q_H^{(3)}$ for the new facilities.

for this problems, and is notably better than the average market share produced by GA after 5000 function evaluations.

### 4.3 Probability to Find Optimal Solution

The experimental investigation showed, that all four heuristic algorithms investigated in this paper are able to find the optimal solution, its determination is not guaranteed. Therefore it is important to investigate the probability to achieve the optimal solution or its approximation with known discrepancy. It was evaluated by cumulative distribution function (CDF), considering the error of approximation of the optimal solution as a real-valued random variable (see [12] for details).

The obtained CDF curves are presented in Figs. 6 and 7, where the horizontal axis stands for the error of the approximation of the optimal solution (in percents of the objective function value), and the vertical axis – for the probability to get the approximation with the corresponding accuracy.

Figure 6 illustrates that RDOA-D-PreMS almost always determines the optimal solution meanwhile GA demonstrate good rate of finding approximation of the optimal solution with 3–4 percents discrepancy after 5000 function evaluations. Figure 7 illustrates lower probability to find the optimal solution when qualities of the new facilities are high: the probability to find the optimal solution using RDOA-D-PreMS is around 0.4, meanwhile GA finds the optimal solution with probability around 0.2 and around 0.1 for the instances with 500 and 1000 candidate locations, respectively.

The RDOA-D-PreMS always finds an approximation of the optimal solution with 3% discrepancy for the instance with 500 candidate locations, meanwhile the probability to find an approximation with lower than 5% discrepancy for the instance with 1000 candidate locations is 0.92.

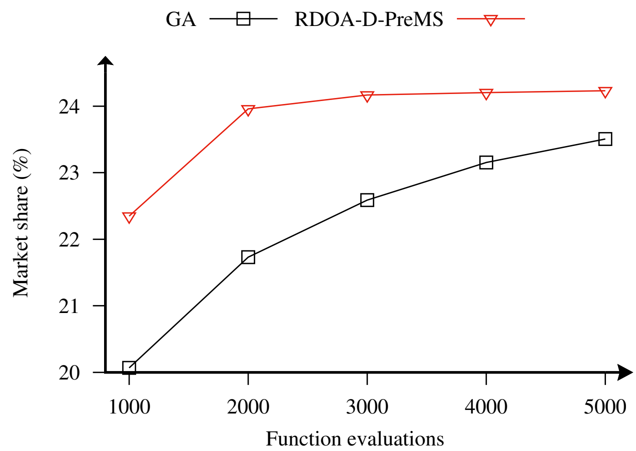


**Fig. 5** Results obtained when choosing locations for 5 new facilities from 500 candidate locations considering low qualities $Q_L^{(5)}$ for the new facilities.

## 5 Conclusions

The competitive facility location problem with binary customer choice rule and asymmetric objective function was formulated as a mixed integer linear programming problem and heuristic strategies to deal with large sets of data were designed. The proposed algorithms use a strategy for ranking of the candidate locations, which was applied for symmetric problems in our previous work, and redesigned to make it suitable for asymmetric problems. Additionally, new algorithms include storing of a set of the best solutions found so far, which

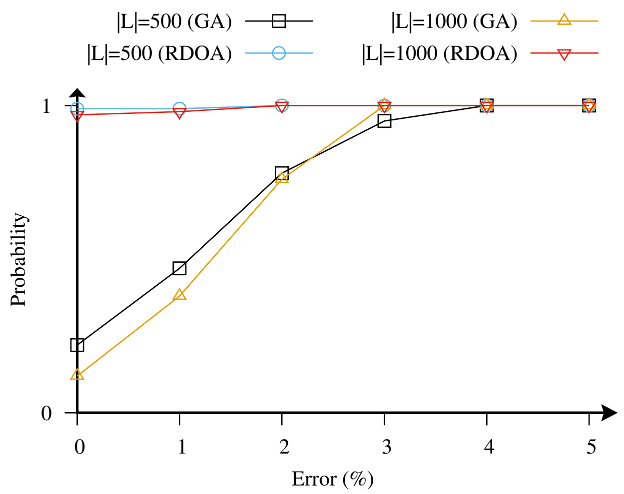


**Fig. 6** Probability to obtain the optimal solution with known discrepancy when choosing locations for 3 new facilities from 500 candidate locations considering low qualities $Q_L^{(3)}$ for the new facilities.

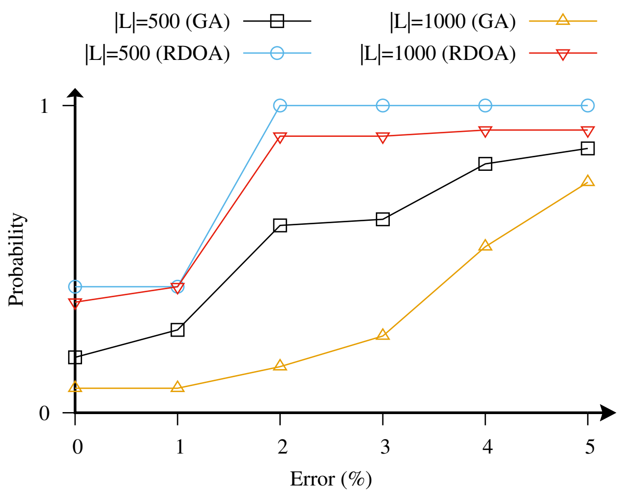


**Fig. 7** Probability to obtain the optimal solution with known discrepancy when choosing locations for 3 new facilities from 500 candidate locations considering high qualities $Q_H^{(3)}$ for the new facilities.

size is reduced automatically, and the evaluation of upper bounds for the market share of the new facilities, which was used to reject obviously unacceptable solutions without evaluation of the objective function.

The results of the experimental investigation demonstrate that the proposed heuristic RDOA-D-PreMS is able to determine the optimal solution of different instances of facility location problems, and notably outperforms Genetic Algorithm, which is considered as a good strategy for such a kind of problems.

**Acknowledgements** This research is funded by the European Social Fund under the No 09.3.3-LMT-K-712 “Development of Competences of Scientists, other Researchers and Students through Practical Research Activities” measure.

This research is funded by the Ministry of Economy and Competitiveness of Spain under the research project MTM2015-70260-P, and by the Fundación Séneca (The Agency of Science and Technology of the Region of Murcia) under the research project 19241/PI/14.

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

**Ethical approval** This article does not contain any studies with human participants or animals performed by any of the authors.

## References

1. Atta, S., Sinha Mahapatra, P.R., Mukhopadhyay, A.: Multi-objective uncapacitated facility location problem with customers’ preferences: Pareto-based and weighted sum ga-based approaches. Soft Computing pp. 1–16 (2019). DOI 10.1007/s00500-019-03774-1
2. Chan, K.Y., Aydin, M.E., Fogarty, T.C.: Main effect fine-tuning of the mutation operator and the neighbourhood function for uncapacitated facility location problems. Soft Computing **10**(11), 1075–1090 (2006). DOI 10.1007/s00500-005-0044-4
3. Church, R., ReVelle, C.: The maximal covering location problem. Papers of Regional Science Association **32**(1), 101–118 (1974)
4. Drezner, T., Drezner, Z.: Finding the optimal solution to the Huff based competitive location model. Computational Management Science **1**(2), 193–208 (2004)
5. Farahani, R.Z., Rezapour, S., Drezner, T., Fallah, S.: Competitive supply chain network design: An overview of classifications, models, solution techniques and applications. Omega **45**(0), 92–118 (2014)
6. Fernandes, D., Rocha, C., Aloise, D., Ribeiro, G., Santos, E., Silva, A.: A simple and effective genetic algorithm for the two-stage capacitated facility location problem. Computers and Industrial Engineering **75**, 200–208 (2014)
7. Fernández, P., Pelegrín, B., Lančinskas, A., Žilinskas, J.: New heuristic algorithms for discrete competitive location problems with binary and partially binary customer behavior. Computers & Operations Research **79**, 12–18 (2017)
8. FICO Xpress Mosel: Fair Isaac Corporation (2014)
9. Francis, R.L., Lowe, T.J., Tamir, A.: Demand point aggregation for location models. In: Z. Drezner, H. Hamacher (eds.) Facility Location: Application and Theory, pp. 207–232. Springer-Verlag (2002)
10. Friesz, T., Miller, T., Tobin, R.: Competitive networks facility location models: a survey. Papers in Regional Science **65**, 47–57 (1998)
11. Hakimi, L.: Location with spatial interactions: Competitive locations and games. In: Z. Drezner (ed.) Facility

Location: A Survey of Applications and Methods, pp. 367–386. Springer (1995)
12. Hendrix, E., Lančinskas, A.: On benchmarking stochastic global optimization algorithms. Informatica **26**(4), 649–662 (2015)
13. Holland, J.H.: Adaptation in Natural and Artificial Systems. The University of Michigan Press (1975)
14. Huff, D.L.: Defining and estimating a trade area. Journal of Marketing **28**, 34–38 (1964)
15. Lančinskas, A., Fernández, P., Pelegín, B., Žilinskas, J.: Improving solution of discrete competitive facility location problems. Optimization Letters **11**(2), 259–270 (2017)
16. Onwubolu, G., Mutingi, M.: A genetic algorithm approach to cellular manufacturing systems. Computers and Industrial Engineering **39**(1), 125–144 (2001)
17. Peeters, P.H., Plastria, F.: Discretization results for the Huff and Pareto-Huff competitive location models on networks. TOP **6**, 247–260 (1998)
18. Peng, X., Xia, X., Zhu, R., Lin, L., Gao, H., He, P.: A comparative performance analysis of evolutionary algorithms on k-median and facility location problems. Soft Computing **22**(23), 7787–7796 (2018). DOI 10.1007/s00500-018-3462-9
19. Plastria, F.: Static competitive facility location: An overview of optimisation approaches. European Journal of Operational Research **129**(3), 461–470 (2001)
20. Reeves, C., Rowe, J.: Genetic Algorithms: Principles and Perspectives: A Guide to GA Theory. Kluwer Academic Publishers (2002)
21. ReVelle, C., Eiselt, H., Daskin, M.: A bibliography for some fundamental problem categories in discrete location science. European Journal of Operational Research **184**(3), 817–848 (2008)
22. Serra, D., Colomé, R.: Consumer choice and optimal locations models: Formulations and heuristics. Papers in Regional Science **80**(4), 439–464 (2001)
23. Serra, D., ReVelle, C.: Competitive location in discrete space. In: Z. Drezner (ed.) Facility Location: A Survey of Applications and Methods, pp. 367–386. Springer (1995)
24. Suárez-Vega, R., Santos-Penate, D.R., Dorta-Gonzalez, P.: Discretization and resolution of the $(r|X_p)$-medianoid problem involving quality criteria. TOP **12**(1), 111–133 (2004)
25. Suárez-Vega, R., Santos-Penate, D.R., Dorta-González, P.: The follower location problem with attraction thresholds. Papers in Regional Science **86**(1), 123–137 (2007)
26. Watanabe, M., Ida, K., Gen, M.: A genetic algorithm with modified crossover operator and search area adaptation for the job-shop scheduling problem. Computers and Industrial Engineering **48**(4), 743–752 (2005)