Journal of Computational and Applied Mathematics xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of Computational and Applied Mathematics

journal homepage: www.elsevier.com/locate/cam

Exact and heuristic solutions of a discrete competitive location model with Pareto-Huff customer choice rule

Pascual Fernández [a], Blas Pelegrín [a], Algirdas Lančinskas [b], Julius Žilinskas [b,*]

[a] Department of Statistics and Operations Research, University of Murcia, Spain
[b] Institute of Data Science and Digital Technologies, Vilnius University, Akademijos 4, LT-08663 Vilnius, Lithuania

ARTICLE INFO

Article history:
Received 25 November 2019
Received in revised form 20 August 2020

Keywords:
Competitive location
Pareto-Huff customer choice rule
Ranking-based optimization algorithm

ABSTRACT

An entering firm wants to compete for market share of an area by opening some new facilities selected among a finite set of potential locations (discrete space). Customers are spatially separated and there already are other firms operating in that area. In this paper, we use a variant of the well known Huff (proportional) customer choice rule, the so called Pareto-Huff, which have had little attention on the literature because of its nonlinear formulation. This untested rule considers that customers split their demand among the facilities that are Pareto optimal with respect to quality (to be maximized) and distance (to be minimized), proportionally to their attractions, i.e., a distant facility will capture demand of a customer only if it has higher quality than any other closer facility. A first formulation as a nonlinear programming problem is proposed, and then an equivalent formulation as a linear programming problem is presented, which allows us to obtain exact solutions for medium size problems. For large size problems, a heuristic procedure is also proposed to obtain the best approximate solutions. Its performance is checked for small size problems and its solutions are compared with the optimal solutions given by a standard optimizer, Xpress, using real geographical coordinates and population data of municipalities in Spain.

1. Introduction

When a new firm wants to compete for market share in a given geographical area, one of the most important decisions is where to locate its facilities. Depending on location space, facility attraction, customer behavior and demand function, different location models and solution procedures have been proposed (see [1–3]). The entering firm has to decide the locations for the new facilities in order to maximize its market share or profit, but taking into account the customers' behavior. Traditionally, it was assumed that customers chose the nearest facility to be served, but on real problems, customers take into account some other characteristics of the facilities. Huff [4] proposed the attraction model, where the attraction of a facility is defined as the quotient between facility quality (it depends on its characteristics) and a non-negative non-descending function of the distance between the customer and the facility. The two most common customer choice rules are the Huff (proportional or probabilistic) and the binary (deterministic) rules (see [5]). In the first one, customers patronize all the facilities in proportion to facility attraction (see for instance [6,7]), and in the second one, each customer patronizes only one facility, the one with maximum attraction (see [8,9]). In this paper, we are interested in using a customer choice rule that fits the customers behavior best when they are concentrated in demand points. In

* Corresponding author.
E-mail address: julius.zilinskas@mii.vu.lt (J. Žilinskas).

https://doi.org/10.1016/j.cam.2020.113200

this way, we are going to consider a modification of the Huff rule, the so called Pareto-Huff customer choice rule [10], in which a customer will patronize a more distant facility only if it has higher quality than any other closer facility. Then, each customer will split its demand among the facilities that are Pareto optimal with respect to quality (to be maximized) and distance (to be minimized), proportionally with their attraction.

The Pareto-Huff model has been formulated as a nonlinear programming problem [10], and to our knowledge, no linear formulation has been proposed in the literature. In this paper we propose a linearization of the Pareto-Huff model as a binary linear programming problem, so this model can be solved exactly by using standard optimization software, at least for medium size data. For greater size data, it is shown that the heuristic algorithm that we proposed in Fernández et al. (2017) can be updated using new sampling probabilities obtaining excellent results for this new model.

The remainder of the paper is organized as follows: Section 2 consists of description of the competitive location problem and its formulation as both, nonlinear and linearized problems. Section 3 is devoted to presentation of the heuristic algorithm, and Section 4 includes the description and discussion of the experimental investigation. Finally, conclusions are presented in Section 5.

2. Discrete location models

Consider a given area where customers are supposed to be aggregated to geographic demand points (see [11] for demand aggregation). Their demands are fixed and known. Different facilities belonging to different firms are already located in that region. An entering firm wants to open new facilities in order to capture as much demand as possible, taking into account the pre-existing facilities already located. We will assume for simplicity that all pre-existing facilities belong to the same firm, the competitor.

The following general notation is used:

Indices:

i, I	index and set of demand points (customers)
j, h, k	indices of location facilities

Data:

w_i	demand at i.
q_j	quality at location j.
d_{ij}	distance between demand point i and location j.
a_{ij}	attraction that demand point i feels for a new facility at location j.
$a_i(S)$	maximum attraction that i feels for facilities in the set S $a_i(S) = \max\{a_{ij} : j \in S\}$
L	set of candidate locations for the new facilities.
C	set of pre-existing facilities of competitors.
s	number of new facilities to be located.

Variables:

X	set of locations for the new facilities.

In this new model, the demand of each customer i will not be split between all open facilities, but only among facilities that are Pareto optimal with respect to quality and distance. The set of such facilities is denoted by PH_i. Demand will be split proportionally with the attraction that customer feels to facilities in PH_i. A customer will patronize a more distant facility only if it has higher quality than any other closer facility, so a distant facility will be selected by a customer only if no facilities exist that are both closer and at least with the same quality. Facilities belonging to PH_i are non-dominated facilities with respect to quality and distance, i.e., for any $j \in PH_i$ there does not exist any other facility with the same quality and closer to i than j, or to the same distance to i and with a quality greater than q_j (see Fig. 1).

If $M_{PH}(X)$ denotes the market share captured by the entering firm when its new facilities are located at X, the Pareto-Huff problem can be formulated as:

$$Max\{M_{PH}(X) = \sum_{i \in I} w_i \frac{\sum_{j \in PH_i \cap X} a_{ij}}{\sum_{j \in PH_i} a_{ij}} : |X| = s, X \subset L\} \tag{1}$$

which is a nonlinear optimization problem.

In order to linearize this model, the following order relationship on the set of facilities is defined for each customer i:

$$k \succ^i j \sim \begin{cases} d_{ik} < d_{ij}, q_k \geq q_j \\ \quad or \\ d_{ik} = d_{ij}, q_k > q_j \end{cases} \tag{2}$$

which indicates that, for a given consumer i, if a facility k dominates another facility j, then j will not be a facility to serve demand point i, so $PH_i = \{j \in C \cup X : \nexists k \in C \cup X \text{ such that } k \succ^i j\}$.

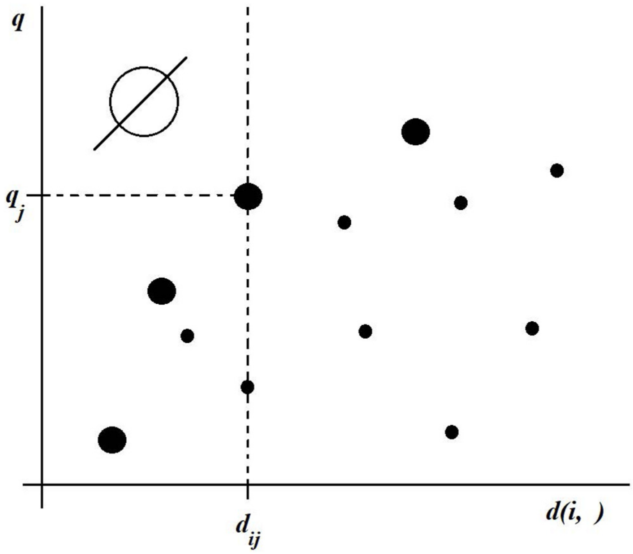

Fig. 1. PH_i is represented by the big circles.

Taking into account this dominance relationship and knowing the set of pre-existing facilities, for each customer i, it is possible to define the subsets of competitors' facilities (C_i) and candidate locations (L_i) that can be Pareto optimal facilities, since they are not dominated by any pre-existing facility:

$$\begin{aligned} C_i &= \{h \in C : \nexists k \in C \text{ such that } k \succ^i h\} \\ L_i &= \{j \in L : \nexists k \in C \text{ such that } k \succ^i j\} \end{aligned} \tag{3}$$

To propose a formulation of the model, the following binary variables are considered:

$$x_j = \begin{cases} 1 & \text{if a new facility is located at } j \\ 0 & \text{otherwise} \end{cases} \qquad j \in L$$

$$y_{ij} = \begin{cases} 1 & \text{if } i \text{ is partially served by a new facility } j \\ 0 & \text{otherwise} \end{cases} \qquad i \in I, j \in L_i$$

$$z_{ih} = \begin{cases} 1 & \text{if } i \text{ is partially served by an existing facility } h \\ 0 & \text{otherwise} \end{cases} \qquad i \in I, h \in C_i$$

So, the objective function of this model is:

$$M_{PH}(X) = \sum_{i \in I} w_i \frac{\sum_{j \in PH_i \cap X} a_{ij}}{\sum_{j \in PH_i} a_{ij}} = \sum_{i \in I} w_i \frac{\sum_{j \in L_i} a_{ij} y_{ij}}{\sum_{j \in L_i} a_{ij} y_{ij} + \sum_{h \in C_i} a_{ih} z_{ih}} \tag{4}$$

The variables of the model must verify that:

(i) $z_{ih} = 1$ if $h \in C_i \cap PH_i$
(ii) $y_{ij} = 1$ if $x_j = 1$ and $j \in L_i \cap PH_i$

To formulate the constraints that ensure the above conditions hold, we define the following sets:

$$\begin{aligned} D_{ih}^C &= \{k \in L_i : k \succ^i h\}, \forall i \in I, \forall h \in C_i \\ D_{ij}^L &= \{k \in L_i : k \succ^i j\}, \forall i \in I, \forall j \in L_i \end{aligned} \tag{5}$$

Then, condition (i) is equivalent to:

$$z_{ih} = 1 \Leftrightarrow x_k = 0, \forall k \in D_{ih}^C \tag{6}$$

for which the following constraint sets are required:

$$\begin{aligned} &\sum_{k\in D^C_{ih}} x_k + z_{ih} \geq 1, \forall i \in I, \forall h \in C_i \\ &\sum_{k\in D^C_{ih}} x_k \leq |L_i|(1 - z_{ih}), \forall i \in I, \forall h \in C_i \end{aligned} \tag{7}$$

On the other hand, condition (ii) is equivalent to:

$$y_{ij} = 1 \Leftrightarrow x_j = 1 \text{ and } x_k = 0, \forall k \in D^L_{ij} \tag{8}$$

and this condition would hold if the following constraint sets are considered:

$$\begin{aligned} &y_{ij} \leq x_j, \forall i \in I, \forall j \in L_i \\ &\sum_{k\in D^L_{ij}} x_k + y_{ij} \geq x_j, \forall i \in I, \forall j \in L_i \\ &\sum_{k\in D^L_{ij}} x_k \leq |L_i|(1 - y_{ij}), \forall i \in I, \forall j \in L_i \end{aligned} \tag{9}$$

Then, the proposed problem has the following formulation:

$$\left\{ \begin{array}{ll} \text{Max} & \displaystyle\sum_{i\in I} w_i \frac{\displaystyle\sum_{j\in L_i} a_{ij}y_{ij}}{\displaystyle\sum_{j\in L_i} a_{ij}y_{ij} + \sum_{h\in C_i} a_{ih}z_{ih}} \\ \text{s.t.} & \displaystyle\sum_{j\in L} x_j = s \\ & \displaystyle\sum_{k\in D^C_{ih}} x_k + z_{ih} \geq 1, \forall i \in I, \forall h \in C_i \\ & \displaystyle\sum_{k\in D^C_{ih}} x_k \leq |L_i|(1 - z_{ih}), \forall i \in I, \forall h \in C_i \\ & y_{ij} \leq x_j, \forall i \in I, \forall j \in L_i \\ & \displaystyle\sum_{k\in D^L_{ij}} x_k + y_{ij} \geq x_j, \forall i \in I, \forall j \in L_i \\ & \displaystyle\sum_{k\in D^L_{ij}} x_k \leq |L_i|(1 - y_{ij}), \forall i \in I, \forall j \in L_i \\ & x_j \in \{0, 1\}, \forall j \in L \\ & y_{ij} \in \{0, 1\}, \forall i \in I, \forall j \in L_i \\ & z_{ih} \in \{0, 1\}, \forall i \in I, \forall h \in C_i \end{array} \right. \tag{10}$$

which is a nonlinear binary programming problem.

2.1. Linearization of the pareto-huff model

The formulation of Pareto-Huff model proposed in the previous Section is nonlinear due to the objective function, but it can be linearized in two steps. In the first step, by using the following set of variables:

$$u_i = \frac{1}{\displaystyle\sum_{j\in L_i} a_{ij}y_{ij} + \sum_{h\in C_i} a_{ih}z_{ih}}, \forall i \in I \tag{11}$$

and adding a new constraint set:

$$\sum_{j\in L_i} a_{ij}u_iy_{ij} + \sum_{h\in C_i} a_{ih}u_iz_{ih} = 1, \forall i \in I \tag{12}$$

and in the second step, by using two new variable sets defined as the product of continuous and binary variables, and its corresponding constraint sets:

New variables	Indices	Constraints
$v1_{ij} = u_i y_{ij}$	$\forall i \in I$	$v1_{ij} \le M y_{ij}$
	$\forall j \in L_i$	$v1_{ij} \le u_i$
		$u_i \le v1_{ij} + M(1 - y_{ij})$
		$v1_{ij} \ge 0$
$v2_{ih} = u_i z_{ih}$	$\forall i \in I$	$v2_{ih} \le M z_{ih}$
	$\forall h \in C_i$	$v2_{ih} \le u_i$
		$u_i \le v2_{ih} + M(1 - z_{ih})$
		$v2_{ih} \ge 0$

(13)

where $M = \max\{\frac{1}{a_{ij}} : i \in I, j \in L_i \cup C_i\}$ (see [12–15]).

Then, the discrete competitive facility location model with Pareto-Huff customer choice rule has the following formulation as a mixed binary linear programming problem:

$$\left\{\begin{array}{lll}
\text{Max} & \displaystyle\sum_{i\in I}\sum_{j\in L_i} w_i a_{ij} v1_{ij} & \\
\text{s.t.} & \displaystyle\sum_{j\in L} x_j = s,\ x_j \in \{0,1\}, \forall j \in L & \\
& y_{ij} \le x_j, \forall i, \forall j & v1_{ij} \le M y_{ij}, \forall i, \forall j \\
& \displaystyle\sum_{k\in D^C_{ih}} x_k + z_{ih} \ge 1, \forall i, \forall h & v1_{ij} \le u_i, \forall i, \forall j \\
& \displaystyle\sum_{k\in D^C_{ih}} x_k \le |L_i|(1 - z_{ih}), \forall i, \forall h & u_i \le v1_{ij} + M(1 - y_{ij}), \forall i, \forall j \\
& \displaystyle\sum_{k\in D^L_{ij}} x_k + y_{ij} \ge x_j, \forall i, \forall j & v2_{ih} \le M z_{ih}, \forall i, \forall h \\
& \displaystyle\sum_{k\in D^L_{ij}} x_k \le |L_i|(1 - y_{ij}), \forall i, \forall j & v2_{ih} \le u_i, \forall i, \forall h \\
& \displaystyle\sum_{j\in L_i} a_{ij} v1_{ij} + \sum_{h\in C_i} a_{ih} v2_{ih} \le 1, \forall i & u_i \le v2_{ih} + M(1 - z_{ih}), \forall i, \forall h \\
& u_i \ge 0, \forall i \in I & \\
& y_{ij} \in \{0,1\}, \forall i, \forall j & v1_{ij} \ge 0, \forall i, \forall j \\
& z_{ih} \in \{0,1\}, \forall i, \forall h & v2_{ih} \ge 0, \forall i, \forall h \\
& (\text{where } i \in I, j \in L_i, h \in C_i) &
\end{array}\right.$$

3. Ranking-based discrete optimization algorithm

The Ranking-based Discrete Optimization Algorithm (RDOA) starts from a randomly generated initial solution

$$X = \{x_1, x_2, \dots, x_s\} \tag{14}$$

where s is the number of facilities expected to locate. The solution X is a subset of candidate locations L and is considered as the best solution found so far. A new solution

$$X' = \{x'_1, x'_2, \dots, x'_s\} \tag{15}$$

is derived from X by changing some locations for the new facilities. Each locations x_i has probability $1/s$ to be changed and inverse probability — to be copied without change. In case of change, a new location is randomly sampled from the set L of all candidate locations excluding those which already forms X or X':

$$x'_i = \begin{cases} l \in L \setminus (X \cup X'), & \text{if } \xi_i < 1/s, \\ x_i, & \text{otherwise} \end{cases} \tag{16}$$

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

Each candidate location $l_i \in L$ has a rank value r_i which is expressed by a positive integer value and defines the fitness of l_i to form a new solution. At the beginning $r_i = 1$ and it is dynamically adjusted depending on successes and failures when selecting l_i to form a new solution X'. If the market share $M(X')$ captured by the new solution is greater than the market share $M(X)$ captured by the best known solution, then (1) the ranks of all locations which form X' are increased by one and (2) the ranks of all locations that form outperformed solution X, but do not form X' are reduced by one:

$$r_i = \begin{cases} r_i + 1, & \text{if } l_i \in X', \\ r_i - 1, & \text{if } l_i \in X \setminus X', \\ r_i, & \text{otherwise} \end{cases} \tag{17}$$

If $M(X')$ is not greater than $M(X)$, then the ranks of all candidate locations forming unsuccessfully generated solution X', but which do not form the best known solution X, are reduced by one:

$$r_i = \begin{cases} r_i - 1, & \text{if } l_i \in X' \setminus X, \\ r_i, & \text{otherwise} \end{cases} \tag{18}$$

If a rank value becomes equal to zero, then all ranks are increased by one to avoid zero ranks.

The rank values are used to define a probability π_i to sample a candidate location l_i with a rank value r_i to form a new candidate location in (16): the larger rank — the larger sampling probability. This research uses three expressions of the sampling probability, which in addition to the rank value includes other features of the competitive facility location problem.

The first sampling probability expression is based on the composition of the rank r_i and geographical distance between the candidate location l_i and the location $x_k \in X$, which is subject to change (see (16)):

$$\pi_i^{rd} = \frac{r_i}{d(l_i, x_k) \sum_{j=1}^{|L|} \frac{r_j}{d(l_j, x_k)}} \tag{19}$$

where $d(\cdot, \cdot)$ is the distance between two geographical points. The ranks and the distances are normalized – mapped to the interval [0, 1] – to make them equally important to the sampling probability. The smaller distance and larger rank value means better fitness of a candidate location. This expression was proposed, investigated, and compared with the expression based on ranks only in [13].

In this paper we propose to include quality indicator in sampling probability expression assuming that candidate locations with larger quality indicators should have larger sampling probability. Thus the second expression is based on the composition of the rank r_i and the quality q_i of the location l_i:

$$\pi_i^{rq} = \frac{r_i \cdot q_i}{\sum_{j=1}^{|L|} (r_j \cdot q_j)} \tag{20}$$

where ranks and qualities of facilities are normalized.

The third expression of the sampling probability composes rank, distance, and quality:

$$\pi_i^{rdq} = \frac{r_i \cdot q_i}{d(l_i, x_k) \sum_{j=1}^{|L|} \frac{r_j \cdot q_j}{d(l_j, x_k)}} \tag{21}$$

where ranks, qualities, and distances are normalized. A larger sampling probability is assigned to the candidate location with larger rank and quality values and smaller distance to the location being changed.

If a newly generated solution outperforms the best solution found so far, then X is changed by X' and the iteration is assumed to be successful; otherwise, X remains unchanged and the iteration is assumed to be unsuccessful. The process continues till the predefined number of function evaluations is performed and the best solution found so far is returned as the result.

Depending on the expression of sampling probability, the algorithm is abbreviated by RDOA/RD, RDOA/RQ, and RDOA/RDQ, respectively.

4. Experimental investigation

The proposed algorithm has been experimentally investigated by solving the considered model using real geographical data of coordinates and population of 589 municipalities (which will be considered as demand points and its demand equal to the population) in Spain. The distances between demand points and facilities have been calculated in kilometers using Haversine distance [16], and the attractiveness that the demand point i feels for the facility j has been taken as

$$a_{ij} = \frac{q_j}{1 + d_{ij}}. \tag{22}$$

Due to stochastic nature of the algorithm, each experiment has been performed 100 times and average results were recorded and analyzed.

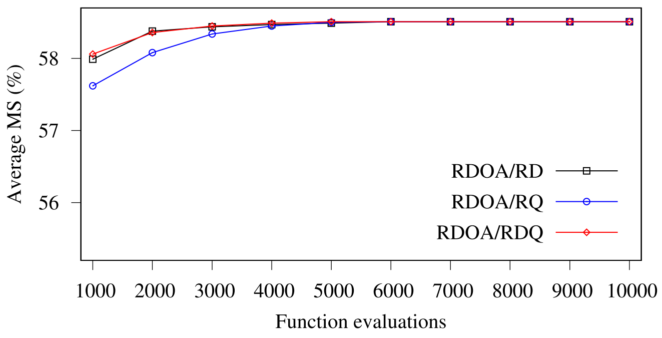

Fig. 2. Performance of the algorithms with different expressions of the sampling probabilities, applied to solve CFLP with 100 candidate locations.

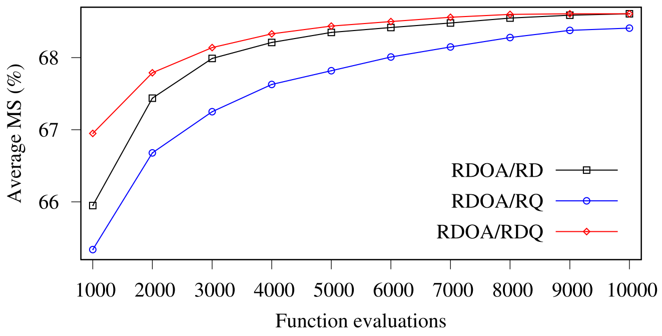

Fig. 3. Performance of the algorithms with different expressions of the sampling probabilities, applied to solve CFLP with 300 candidate locations.

4.1. Impact of the sampling probability

The impact of different expressions of sampling probabilities for candidate locations (see Eqs. (19)–(21)) have been investigated by two instances of the proposed model, which differs on the set L of candidate locations: 100 and 300 candidates. Qualities have been randomly generated over the interval [30, 70]. The set I of 589 demand points representing municipalities in Spain with at least 1000 residents, and the set C of 10 preexisting location with predefined quality values and located in 10 most populated demand points have been used to set up the CFLP.

RDOA/RD, RDOA/RQ, and RDO/RDQ, which differ on the sampling probability expression (see Section 3), have been used to determine the set X of locations for $s = 10$ new facilities. All algorithms have been run for 10,000 function evaluations. Each experiment have been run for 100 times and average results have been recorded at every 1000 function evaluations. The results obtained when solving the CFLP with $|L| = 100$ are presented in Fig. 2, where the horizontal axis represents the number of function evaluations and the vertical axis represents the market share obtained by the new locations, expressed in percents of the total market share in the region.

The results show that change of the distance factor to the quality in the sampling probability expression reduces performance in early stage of the algorithm (compare RDOA/RD with RDOA/RQ in Fig. 2). On the other hand, inclusion of the quality indicator together with the distance slightly improves the performance at the beginning of computations (compare RDOA/RD with RDOA/RQ in Fig. 2), though all algorithms demonstrate the same performance after 5000 function evaluations.

Larger differences in performance of the algorithms have been seen when solving CFLP with the larger set of candidate locations, $|L| = 300$. Results are presented in Fig. 3. One can see from the figure that change of the distance factor to the quality really reduces overall performance of the algorithm (compare RDOA/RD with RDOA/RQ in Fig. 3), but usage of the quality indicator together with the distance notably improves the performance, especially in the early stage of the algorithm (compare RDOA/RD with RDOA/RDQ in Fig. 3).

In general, it is not useful to use the quality indicator instead of the distance, but usage of the quality indicator besides the distance can improve the performance of the algorithm and the improvements are more notable for the larger set of candidate locations.

4.2. Validation of the heuristics

A more extensive investigation of the performance of RDOA/RDQ has been carried out to see the possibility to determine the optimal solution or its approximation. Three sets of preexisting facilities have been used: 10, 20, and

Table 1
Results obtained when solving CFLPs with randomly generated quality values for candidate locations.

$\lvert C\rvert$	$\lvert L\rvert$	s	Xpress		RDOA/RDQ			
			MS	Time (s)	Max	Avg	Std (%)	Error (%)
10	50	5	44.95	708.9	44.95	44.95	0.00	0.00
20	50	5	30.16	74.7	30.16	30.16	0.00	0.00
30	50	5	19.68	22.7	19.68	19.68	0.00	0.00
10	50	10	55.39	2411.3	55.39	55.39	0.00	0.00
20	50	10	44.36	116.6	44.36	44.36	0.00	0.00
30	50	10	31.52	21.0	31.52	31.52	0.00	0.00
10	100	5	45.05	3028.9	45.05	45.05	0.00	0.00
20	100	5	29.96	1441.7	30.16	30.10	0.59	1.96
30	100	5	19.68	89.2	19.68	19.68	0.00	0.00
10	100	10	58.51	7914.8	58.51	58.51	0.00	0.00
20	100	10	45.18	2902.8	45.18	45.08	0.25	0.63
30	100	10	31.87	87.0	31.87	31.87	0.03	0.30
10	200	5	47.63	18001.7	**50.58**	50.58	0.00	0.00
20	200	5	32.50	18003.9	**33.69**	33.51	0.73	1.53
30	200	5	25.97	1931.3	25.97	25.97	0.00	0.00
10	200	10	65.54	18002.0	**67.66**	67.64	0.30	3.04
20	200	10	49.02	18002.9	**50.70**	50.51	0.54	1.66
30	200	10	40.14	18000.9	40.14	40.14	0.08	0.39
10	300	5	48.09	18009.0	**50.97**	50.97	0.00	0.00
20	300	5	30.71	18003.9	**33.69**	33.50	0.81	3.12
30	300	5	25.97	6693.6	25.97	25.97	0.00	0.00
10	300	10	64.64	18016.4	**68.67**	68.61	0.43	2.82
20	300	10	46.57	18004.9	**51.08**	50.93	0.50	1.93
30	300	10	39.76	18001.1	**40.69**	40.48	0.76	2.05

30 facilities, located in the most populated demand points. Their quality values have been randomly generated over the interval [30, 70]. Two numbers of the new facilities have been used: 5 and 10. Four sets of 50, 100, 200, and 300 most populated demand points have been considered as location candidates for the new facilities. Combinations of these parameters create 24 different instances.

The quality values of the new facilities depend on the quality values of candidate locations. Three different scenarios for generation of quality values for the candidate locations have been considered. In the first scenario qualities are random integers from the interval [30, 70], in the second scenario quality value of a candidate location is proportional to its demand, and in the third one quality value of a candidate location is inversely proportional to its demand. If a new facility is located where there is already a preexisting facility but with a higher quality, all the demand for that location will be captured by the new facility. As the model tries to maximize the total demand captured by the new facilities, to avoid co-location in the highest demand points, it has been assumed that the qualities for the new facilities in these points are less than or equal to the qualities of the preexisting ones. This has been taken into account in the second scenario for the 5 or 10 most populated points, depending on s value.

These three scenarios of generation of quality values for location candidates expand the number of instances to 72 (24 instances per scenario). The same set of 589 demand points as in the previous investigation has been used in all cases. All problem instances have been solved by deterministic integer linear programming solver Xpress [17]. Xpress worked till the optimal solution is determined or the time limit of 18,000 s is exceeded.

The same instances have been solved by RDOA/RDQ, which has been run for 10,000 function evaluations for each instance. Each experiment has been run for 100 independent runs and statistical results have been recorded.

The results are presented in Tables 1–3, where numbers of preexisting facilities, candidate locations, and new facilities are presented in the first three columns; next two columns stand for the best market share obtained by Xpress (in percents of the total market share in the region) and computational time; the last four columns present statistics of the results obtained by RDOA/RDQ: the maximal and average of the best market share in percents of the total market share, the standard deviation in percents of the average, and the maximal percentage difference between the minimal and the maximal market share. The maximal market share in bold font indicates instances for which RDOA/RDQ founds better solutions than Xpress within a given time limit (18,000 s).

The results show that, depending on the instance, Xpress requires from less than one second to more that 18,000 s to solve the problem. Meanwhile, RDOA/RDQ performs 10,000 function evaluations within around 1 to around 5 s, depending on the computational time required to evaluate the objective function. Instances with less preexisting and new facilities (e.g., the first instance) require less time, while instances with a larger number of preexisting and new facilities require more computational time. The number of candidate locations does not make impact to the computational time.

The tables show that the heuristic algorithm is able to find the optimal solution or at least its approximation with reasonable accuracy. The algorithm found the optimal solution for all instances, in most cases the optimal solution has been determined in all 100 independent runs. In some runs RDOA/RDQ failed to determine the optimal solution, however the maximal discrepancy from the maximal market share was 3.12%, and below 2% in most of the cases.

Table 2
Results obtained when solving CFLPs with quality values of candidate locations proportional to demand values.

$\|C\|$	$\|L\|$	s	Xpress		RDOA/RDQ			
			MS	Time (s)	Max	Avg	Std (%)	Error (%)
10	50	5	18.47	19.8	18.47	18.47	0.00	0.00
20	50	5	12.49	0.4	12.49	12.49	0.00	0.00
30	50	5	10.73	0.3	10.73	10.73	0.00	0.00
10	50	10	30.09	19.5	30.09	30.09	0.00	0.00
20	50	10	20.14	0.6	20.14	20.14	0.00	0.00
30	50	10	15.96	0.3	15.96	15.96	0.01	0.12
10	100	5	28.45	140.8	28.45	28.43	0.03	0.08
20	100	5	19.99	29.0	19.99	19.99	0.00	0.00
30	100	5	15.70	18.3	15.70	15.70	0.00	0.00
10	100	10	41.96	400.5	41.96	41.96	0.00	0.00
20	100	10	32.86	47.0	32.86	32.86	0.02	0.13
30	100	10	24.10	19.1	24.10	24.10	0.00	0.00
10	200	5	30.92	18003.2	**32.55**	32.48	0.23	0.46
20	200	5	24.00	18000.8	24.04	24.03	0.15	0.53
30	200	5	16.95	1374.7	16.95	16.95	0.02	0.13
10	200	10	44.98	18001.8	**47.42**	47.41	0.20	1.89
20	200	10	38.00	18001.3	**38.98**	38.98	0.11	0.80
30	200	10	28.26	14730.5	28.26	28.25	0.25	1.87
10	300	5	31.49	18001.2	**32.72**	32.59	0.32	0.66
20	300	5	23.50	18012.3	**24.20**	24.19	0.17	0.99
30	300	5	17.08	4007.1	17.08	17.08	0.03	0.13
10	300	10	42.36	18003.4	**47.79**	47.74	0.40	1.68
20	300	10	37.20	18001.1	**39.18**	39.15	0.24	0.87
30	300	10	28.21	18001.9	**28.45**	28.45	0.11	0.47

Table 3
Results obtained when solving CFLPs with quality values of candidate locations inversely proportional to demand values.

$\|C\|$	$\|L\|$	s	Xpress		RDOA/RDQ			
			MS	Time (s)	Max	Avg	Std (%)	Error (%)
10	50	5	33.70	25.6	33.70	33.70	0.00	0.00
20	50	5	21.72	10.4	21.72	21.72	0.00	0.00
30	50	5	14.95	6.3	14.95	14.95	0.00	0.00
10	50	10	40.43	78.3	40.43	40.39	0.08	0.16
20	50	10	27.56	16.8	27.56	27.56	0.00	0.00
30	50	10	21.60	1.5	21.60	21.60	0.00	0.00
10	100	5	34.08	984.4	34.08	34.08	0.00	0.00
20	100	5	22.39	166.6	22.39	22.39	0.00	0.00
30	100	5	13.22	89.2	13.22	13.22	0.17	1.24
10	100	10	46.95	2192.9	46.95	46.95	0.00	0.00
20	100	10	33.92	504.5	33.92	33.88	0.36	1.59
30	100	10	21.91	124.7	21.91	21.82	0.94	2.39
10	200	5	40.75	18004.6	**41.28**	41.28	0.00	0.00
20	200	5	26.11	18002.2	**27.03**	27.03	0.00	0.00
30	200	5	16.06	7424.7	16.06	16.02	0.49	1.28
10	200	10	52.29	18007.4	**54.73**	54.54	0.46	1.23
20	200	10	39.16	18001.4	**39.99**	39.99	0.11	0.82
30	200	10	26.74	18001.0	26.74	26.72	0.21	1.87
10	300	5	41.40	18008.0	**42.42**	42.42	0.00	0.00
20	300	5	29.47	18008.5	29.47	29.43	0.48	1.58
30	300	5	20.56	1480.7	20.56	20.56	0.00	0.00
10	300	10	52.67	18004.9	**54.40**	54.39	0.04	0.14
20	300	10	39.52	18001.5	**41.75**	41.66	0.56	2.67
30	300	10	28.55	18002.3	**29.29**	29.22	0.48	2.81

The most important parameters of the problem are the number of candidate locations and the number of new facilities, since their increment make the problem more complicated for RDOA/RDQ. The number of preexisting facilities make change to the computational time, however it has not been observed that it makes impact for complexity of the problem.

The heuristic algorithm has been used to solve more complicated instances of the problem with 1000 demand points all of which are considered as candidate locations. The instance with 10 preexisting facilities, 10 new facilities, and 1000 candidate locations with randomly generated quality values have been used to investigate the performance of RDOA with different expressions of the sampling probability: RDOA/RD, RDOA/RQ, and RDOA/RDQ. The performance of the algorithms

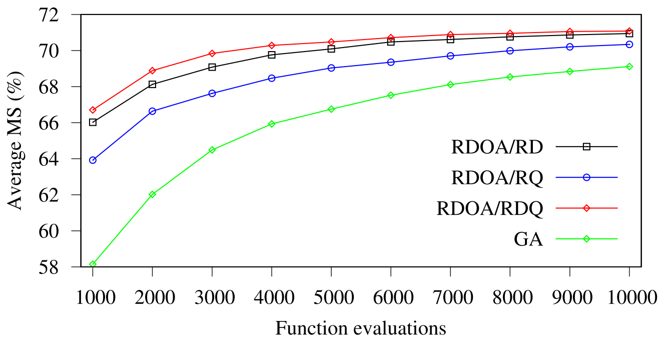

Fig. 4. The performance of RDOA with different sampling probabilities in comparison with performance of GA.

has been compared with the performance of the Genetic Algorithm [18–20], which has been successfully applied to solve DCFLPs in [21]. All four algorithms have the same budget for computational resources — 10,000 function evaluations.

The results are presented in Fig. 4, where the horizontal axis stands for the number of function evaluations, the vertical axis – for the average market share, and different curves – for the performance of different algorithms. One can see from the figure, all three versions of RDOA notably outperform GA. Inclusion of quality indicator instead of geographical distance in expression of sampling probability reduces performance of the algorithm (RDOA/RD is better than RDOA/RQ). The best performance has been achieved when all three features (the rank, the distance, and the quality) are included in the expression of sampling probability.

5. Conclusions

A novel discrete competitive facility location model has been presented where an entering firm wants to locate a fixed number of new facilities selected from a location candidate set in order to maximize the market share when the Pareto-Huff customer choice rule is considered.

An initial formulation of this model as a binary nonlinear programming problem is proposed, which is linearized as a mixed binary linear programming problem. So, the proposed model is finally formulated as a binary linear programming problem. This allows us to solve exactly medium size problems, so a ranking-based-search heuristic algorithm is applied with three different sampling strategies for larger size problems. These strategies have been compared and computational studies show that it is not useful to use the quality indicator instead of the distance, but the usage of the quality indicator in addition to the distance can improve performance of the algorithm and the improvements are more notable when the data size increases.

The best results are obtained by using the RDOA/RDQ method, which uses ranks, qualities and distances to define the sampling probability of each location candidate. The algorithm found the best known solution for all instances, in most cases the best known solution have been determined in all 100 independent runs. In some runs RDOA/RDQ failed to determine the best known solution, however the maximal discrepancy from the maximal market share was 3.12%, and below 2% in most of the cases.

Acknowledgments

This research has been supported by Fundación Séneca (The Agency of Science and Technology of the Region of Murcia, Spain) under the research project 20817/PI/18. The research work of A. Lančinskas was funded by a Grant (No. S-MIP-17-67) from the Research Council of Lithuania.

References

[1] R.Z. Farahani, S. Rezapour, T. Drezner, S. Fallah, Competitive supply chain network design: an overview of classifications, models, solution techniques and applications, Omega 45 (2014) 92–118.
[2] T. Friesz, Competitive networks facility location models: a survey, Pap. Reg. Sci. 65 (1998) 47–57.
[3] C. ReVelle, H. Eiselt, M. Daskin, A bibliography for some fundamental problem categories in discrete location science, European J. Oper. Res. 184 (3) (2008) 817–848.
[4] D. Huff, Defining and estimating a trade area, J. Mark. 28 (1964) 34–38.
[5] D. Serra, H.A. Eiselt, G. Laporte, C.S. ReVelle, Market capture models under various customer-choice rules, Environ. Plan. B: Plann. Des. 26 (1999) 741–750.
[6] T. Drezner, Z. Drezner, Finding the optimal solution to the huff based competitive location model, Comput. Manag. Sci. 1 (2) (2004) 193–208.
[7] D. Serra, R. Colomé, Consumer choice and optimal location models: formulations and heuristics, Pap. Reg. Sci. 80 (4) (2001) 439–464.
[8] L. Hakimi, Location with spatial interactions: competitive locations and games, in: Z. Drezner (Ed.), Facility Location: A Survey of Applications and Methods, 1995, pp. 367–386.

[9] C.M. Campos, D.R. Santos-Penate, J.A. Moreno, An exact procedure and LP formulations for the leader-follower location problem, TOP 18 (1) (2010) 97–121.
[10] P.H. Peeters, F. Plastria, Discretization results for the Huff and Pareto–Huff competitive location models on networks, TOP 6 (1998) 247–260.
[11] R.L. Francis, T.J. Lowe, A. Tamir, Demand point aggregation for location models, in: Z. Drezner, H. Hamacher (Eds.), Facility Location: Application and Theory, 2002, pp. 207–232.
[12] B. Biesinger, B. Hu, G. Raidl, Models and algorithms for competitive facility location problems with different customer behaviour, Ann. Math. Artif. Intell. (2014) http://dx.doi.org/10.1007/s10472-014-9448-0.
[13] P. Fernández, B. Pelegrín, A. Lančinskas, J. Žilinskas, New heuristic algorithms for discrete competitive location problems with binary and partially binary customer behavior, Comput. Oper. Res. 79 (2017) 12–18.
[14] K. Haase, S. Muller, A comparison of linear reformulations for multinomial logit choice probabilities in facility location models, European J. Oper. Res. 232 (2014) 689–691.
[15] Y. Kochetov, N. Kochetova, A. Plyasunov, A matheuristic for the leader-follower facility location and design problem, in: H. Lau, P. Van Hentenryck, G. Raidl (Eds.), Proceedings of the 10th Metaheuristics International Conference, MIC 2013, 2013, 32/132/3.
[16] R.W. Sinnott, Virtues of the haversine, Sky Telesc. 68 (2) (1984) 159.
[17] FICO Xpress Mosel, Fair Isaac Corporation, 2014.
[18] J.H. Holland, Adaptation in Natural and Artificial Systems, University of Michigan Press, Ann Arbor, 1975.
[19] David E. Goldberg, Genetic algorithms in search, optimization and machine learning, 1st, Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, ISBN: 0201157675, 1989.
[20] L. Davis, Handbook of Genetic Algorithms, Van Nostrand Reinhold, New York, 1991.
[21] A. Lančinskas, P. Fernández, B. Pelegrín, J. Žilinskas, Improving solution of discrete competitive facility location problems, Optim. Lett. 11 (2017) 259–270.