

Solving Large Immobile Location-Allocation by Affinity Propagation and Simulated Annealing. Application to Select which Sporting Event to Watch

F. Torrent-Fontbona^{a,*}, V. Muñoz^b, B. López^a

^a*eXiT research group, Institute of Informatics and Applications, University of Girona, Campus Montilivi, Building P4, 17071 Girona*

^b*Newronia S.L., Parc Científic i Tecnològic, Building Narcís Monturiol, 17071 Girona*

Abstract

Immobile Location-Allocation (ILA) is a combinatorial problem which consists in, given a set of facilities and a set of demand points, determining the optimal service each facility has to offer and allocating the demand to such facilities. The applicability of optimization methods is tied up to the dimensionality of the problem, but since the distance between data points is a key factor, clustering techniques to partition the data space can be applied, converting the large initial problem into several simpler ILA problems that can be solved separately. This paper presents a novel method that combines clustering and heuristic methods to solve an ILA problem, which reduces the elapsed time keeping the quality of the solution found compared with other heuristics methods.

Keywords: Clustering, immobile location-allocation, heuristics, affinity propagation, simulated annealing

1. Introduction

Immobile Location-Allocation (ILA), (Pasandideh & Niaki, 2010), is a class of Location-Allocation (LA) problem, (Cooper, 1963). LA is a two step problem where given a set of facilities with unknown positions, and a set of demand points of the service provided by the facility, it consists in (i) determining the optimal location of each facility and (ii) allocating the demand to such facilities, usually to the closest ones. ILA considers a set of services instead of one, so the first step of the problem consists on determining the optimal service each facility has to offer, instead of its location. It is a k_n combinatorial problem where k is the number of services and n the number of facilities. Therefore, the applicability of optimization methods is tied up to the dimensionality of the problem.

ILA (and LA) has been mainly applied with limited number of facilities inside an industry or company. Small and medium enterprises (SME), however, with usually a low service capacity, can take advantage of

information and communication technologies to coordinate their services so as to maximize their capacity while competing for the demand. In this new scenario, the number of facilities to be handled in the ILA problem could be huge, as well as the number of services, and the current off-the-shelf ILA problem solving techniques could become obsolete. This has been our case in the following motivating problem we are posed on: the barmen decision problem regarding which sporting event display in their business. Nowadays a vast number of matches of different sports are played in the same day, and a lot of them are even played simultaneously. All that matches are also broadcast by different TV channels due to the current great interest on sport. Most of them are payment TV channels and a lot of people go to bars to watch the match they want. Globalization involves a demand diversification what means that sometimes customers of the same bar want to watch different matches drawing the barman into a decision problem. An evidence of this is the showing up of customer oriented applications that apprise users from which bars around their location broadcast the match they want to watch. Nevertheless, barmen are still deciding the match they display according their estimation of the audience of each match without taking into account other bars in the region. If all the bars display the most popular match, they will share the same customers while wasting the opportunity to have some business for the customers in-

*Corresponding author

Email addresses: ferran.torrent@udg.edu (F. Torrent-Fontbona), victor.munoz@newronia.com (V. Muñoz), beatriz.lopez@udg.edu (B. López)

URL: <http://eia.udg.es/~ftorrent/> (F. Torrent-Fontbona), www.newronia.com (V. Muñoz), <http://eia.udg.es/~blopez/> (B. López)

interested on other matches. Thus, coordinating the decision of which match to display could redound in an increase of the global benefit. While this situation is repeated, and a fairness mechanism is taken into account, the revenue could be increased in an egalitarian basis for all of the bars participating in the coordinated decision.

In this paper we present a mathematical formulation of the ILA problem with such features: huge number of facilities with a limited capacity but with the capability of dynamically changing the service they provide in repeated problem solving episodes. Since the dimensionality of the ILA problem makes the current solving techniques inapplicable, we provide a novel method to solve the problem that consists in using clustering to divide the problem into subproblems. With clustering, the computational complexity of the problem is reduced and it is possible to apply optimization algorithms to find a good solution to each subproblem.

The method we present here has been experimentally tested in the barmen decision problem, demonstrating the benefits of using our method regarding isolated decisions, as well as the benefits induced by the clustering techniques. Other possible applications include the energy distribution over a vast region, where clustering can be used to divide this region, and locally solve the energy distribution problem of each sub-region while taking into account local renewable generation sources. Water supply can also be a subject of this formulation, as well as other applications, as much as the assumptions of the problem formulation hold.

The paper is organized as follows. First we present related work to contextualize this paper. Second we formalize the problem giving its mathematical formulation. Next, we expound the method we propose, the experimental set up we have used and then we present the results obtained with our method providing a discussion. Finally we expound the conclusions of this work and some future work related to it.

2. Related Work

In (Aboolian et al., 2008, 2009) a model to formulate a location-allocation problem is proposed. In the model several facilities are static and LA is used to determine the location of some extra facilities to serve the demand. In these previous works the facilities offered the same service, whereas in this paper we propose a model for an ILA problem where all facilities are located and we have to decide the service each one offers.

In (Pasandideh & Niaki, 2010) a mathematical model for an ILA problem is proposed that consists in locating service units in a pre-established set of immo-

bile server locations to minimize the travel and queue time. Therefore, their problem consists in determining the number of service units of each immobile server, offering each facility the same service. This paper is different from (Pasandideh & Niaki, 2010) in that we have to determine the service each facility has to offer. Moreover, this paper differs from (Aboolian et al., 2008, 2009; Pasandideh & Niaki, 2010) in the dimensionality of the problem since the number of facilities we have to deal with is much higher (they deal with a maximum of 200 facilities and we have to deal with thousands of them). Therefore, the methods proposed to solve the respective problems are different. Hence in (Aboolian et al., 2008; Pasandideh & Niaki, 2010) heuristic methods are proposed while we propose a combination of clustering methods (to reduce the complexity) and heuristic methods (to solve the derived subproblems). In (Aboolian et al., 2009) it is proposed an exact algorithm and an integer programming approach depending on the features of the problem. Due to the dimensionality of the problem we cannot use them but approximate methods.

In (Brimberg et al., 2006) the authors have to deal with large scale continuous location-allocation problem which complexity is very high. Due to this complexity, they examine three decomposition strategies to reduce the complexity of the problem where two of them use clustering to define the subproblems. The clustering techniques explored in this paper are different from the examined in (Brimberg et al., 2006) due the nature of the problem since they are dealing with a continuous domain with a few number of facilities but we are dealing with a discrete domain with a lot of facilities.

In (GU & Wang, 2010) the authors deal with the problem of static and transportation facility location which consists in determining the location of these transportation facilities (like ambulances) considering the static facilities (like hospitals). The authors consider that there is only one type of service (take people from their home to hospitals using the ambulances) which differs from problem we tackle in this paper. Moreover in (GU & Wang, 2010), authors propose an heuristic method to solve the mentioned problem, one step of which is clustering the dataset to reduce the search space. The number of clusters is determined by the number of hospitals, and the hospitals are designated ad hoc as the clustering centroids. In our work we perform clustering over the distribution of the facilities in the geographical area without predefined centroids.

3. Problem Formulation and Modeling

The problem that concerns us is finding the optimal service each facility should offer in order to maximize the total amount of customers and assigning each customer to a close facility, if possible. The first step toward solving the problem is to develop a problem model. For doing so, we have done three assumptions: (i) any customer will not go to a facility which is not offering its desired service and so it cannot be assigned to such facility, (ii) the probability that a customer decides to go to its assigned facility to watch the service depends, mainly, on the customer-facility distance, (iii) when there are some facilities offering the same service, customers prefer to go to the closest facility if it is not full, so they must be assigned to their closest (and not full) facility.

The model requires the following data:

- Services** $s \in [1, N_s]$
In the barmen decision problem, services are the set of simultaneous matches that can be displayed in the bar (one per bar).
- Facilities** $f \in [1, N_f]$
Each facility has a capacity C_i .
- Customers** $c \in [1, N_c]$
Each customer desire a particular service $M_j \in [1, N_s]$.

The ILA problems we are considering have a big number of costumers N_c and facilities N_f .

Moreover, the following working variables are also required:

- $\mathbf{x}^q = \langle x_1^q, x_2^q, \dots, x_{N_f}^q \rangle$ a q th vector of services which represents an assignment (or candidate solution) of services to facilities, where $x_i^q \in [1, N_s]$ is the service the i th facility offers in the q th solution.
- z_{ij} a binary variable which is 1 when the j th customer is assigned to the i th facility.

The square Euclidean distance between the j th customer and the i th facility, d_{ij}^2 , is used to compute distances.

Thus, the ILA problem consists of two steps:

1. Finding the assignment \mathbf{x}^{qmax} that maximizes the sum of customers in each facility, weighting them by $\frac{1}{1+d_{ij}^2}$, according to the following expressions:

$$\mathbf{x}^{qmax} = \arg \max_{\mathbf{x}^q} \{D(\mathbf{x}^q)\} \quad (1)$$

$$D(\mathbf{x}^q) = \sum_{i=1}^{N_f} \sum_{j=1}^{N_c} \frac{z_{ij}^q}{1+d_{ij}^2} \quad (2)$$

subject to

$$\forall_i \sum_{j=1}^{N_c} z_{ij}^q \leq C_i \quad (3)$$

$$\forall_j \sum_{i=1}^{N_f} z_{ij}^q \leq 1 \quad (4)$$

$$x_i^q \neq M_j \rightarrow z_{ij}^q = 0 \quad (5)$$

The weight given to each customer represents the probability that the customer will go to its assigned facility. This factor tries to model customer's behavior since it is unlikely that a customer will finally go to its assigned facility if it is far away from it, but is highly probable that it will go to its assigned facility if it is close to it. Note that we are only considering the Euclidean distance, but other parameters like customer preferences can be easily included in the model using a new distance function that considers these parameters.

The constraints of the problem are that for all facilities the number of customers assigned to them cannot overflow the capacity of each facility, equation (3); customers cannot be assigned to more than one facility, equation (4); and any customer can be assigned to a facility that is not going to offer customer's desired service, equation (5).

Note that the three assumptions are reflected in the model. Assumption (i) is considered in equation (5); assumption (ii) is considered in the target function, equation (2), since all customers are weighted by the distance between them and their facility; finally, assumption (iii) is also reflected in the target function since we are maximizing the number of customers while we weight them by the distance, therefore, we are minimizing the distance between facilities and their customers.

2. Finding the allocation of the customers of the costumers to the facilities, \mathbf{z}^{qmax} , according to the assignment \mathbf{x}^{qmax} .

As the formulation tells, the goal is the maximization of the demand considering all facilities. However, facilities could be not interested in using this system due to it does not guarantee high percentages of load or occupancy for all facilities: it guarantees the maximization of the global occupancy and that facilities do not exceed their capacity. Occupancy will depend on the diversity

of the demand. If all customers demand the same service, facilities will share the demand according to the customer neighborhood. On the other hand, if all customers want a different service, facilities could offer a different service each one and therefore they would not share the demand. In the average, there would be a number of most preferred services. Thus, ILA solution could involve facilities with high occupancy and some others with low occupancy.

As the process is repeated, a fairness mechanism equilibrates the facility occupancy along time. Thus, we propose the fair-ILA model that takes into account the occupancy of facilities when problems are iterated.

In an iterative model, we assume that the number of facilities does not change, while the demand and the available services could eventually change, from a problem t to the next problem $t + 1$. Let be $\mathbf{x}^{q_{max}}(t)$ the solution of the t ILA problem, with $\mathbf{z}^{q_{max}}(t)$ the assignment of customers to facilities in this solution, such that $z_{ij}^{q_{max}}(t) = 1$ if the j th customer is assigned to the i th facility in the solution, and $M_j(t) = x_i^{q_{max}}(t)$, with $x_i^{q_{max}}(t) \in \mathbf{x}^{q_{max}}(t)$.

Moreover, let be $o_i(t)$ the occupancy of the i th facility in the t solution, that is:

$$o_i(t) = \sum_{j=1}^{N_c} z_{ij}^{q_{max}}(t) \quad (6)$$

To introduce fairness, we propose to use the occupancy information of solution $t - 1$ in the next problem formulation t . Thus, the fair-ILA problem is modeled as follows:

$$D(\mathbf{x}^q) = \sum_{i=1}^{N_f} (2 - o_i(t-1)) \sum_{j=1}^{N_c} \frac{z_{ij}^q}{1 + d_{ij}^2} \quad (7)$$

For the sake of simplicity we do remove the iteration index t of the variables, except $o_i(t-1)$, assuming that they all belong to the current problem at time t .

4. The Method

As said in the Introduction, the applicability of optimization algorithms to solve the ILA problem is tied up to the dimensionality of the problem. For a few number of facilities and services, optimization techniques as linear programming, branch and bound search algorithms can be applied (Melo et al., 2009; Torrent-Fontbona, 2012).

When considering a large amount of facilities and customers, such algorithms are not applicable due to their high response time. We can approximate the ILA

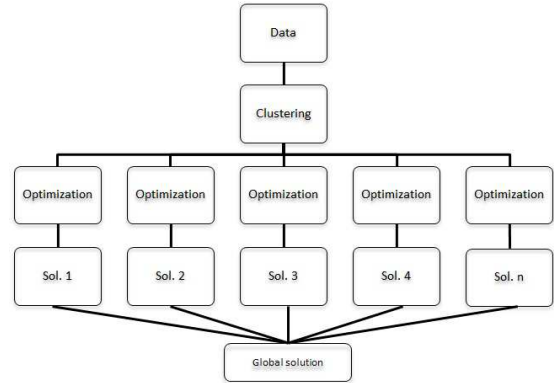


Figure 1: Scheme of the proposed method

solution by dividing the initial problem into several easier subproblems of lower dimension following a divide & conquer strategy. In doing so, we assume that the subproblem solutions do not interact. Although this assumption is strong from an optimization point of view, it enables the application of a method to find a solution. Moreover, one of the ILA components is the geographical distance (location of facilities, distance of customers to the facilities) that enables the division of the problem into subproblems taking into account this space information. Thus, the method we present consists in running a clustering algorithm, which uses a geographical distance, to divide the problem into independent subproblems and then running an optimization algorithm in each subproblem to find a near optimal solution in each one. Finally, the solution to the initial problem is built joining all sub-solutions. Figure 1 illustrates the scheme of the proposed method.

4.1. Clustering

Since the distance between customers and facilities is a key factor, we want to exploit it to simplify the given ILA problem. Thus, the goal of the clustering step is to separate those facilities and customers that are far apart in order to divide the initial problem. For doing so, we aggregate the demand assigning the customers to each closest facility and performing a clustering analysis over the facilities dataset to obtain different clusters of facilities (and customers) and, consequently, dividing the initial problem into different (as much as possible) independent subproblems. Note that, since the target function aims to maximize the total number of assigned customers weighting them by the distance to the facilities, those facilities and customers that are far

away from a region, the influence that they have to the final solution in such region is negligible. Therefore, if clusters are distant enough, subproblems can be considered independent. Thus, there is a trade-off between the simplification of the problem, the number of facilities in each cluster, and how much independent the subproblems are. Then, the clustering algorithm should tend to find small and distant clusters.

The clustering technique we have chosen, after an accurate analysis of the state of the art (Torrent-Fontbona, 2012), is a novel method that has reported good results in different applications called affinity propagation (Frey & Dueck, 2007). This technique performs a clustering analysis that tends to avoid great differences between cluster sizes (with the appropriate initialization) keeping great differences between clusters what is profitable for our purposes. This algorithm, Algorithm 1, considers all data points (facilities) as potential exemplars (data points that correspond to clusters centers) and they exchange two type of real-valued messages in order to "vote" the most representative exemplars. First, each data point sends its "responsibilities" r_{ik} which each one reflects the accumulated evidence for how well-suited point k is to serve as the exemplar for point i , considering other possible exemplars for point i . Second, each point (as a potential exemplar) sends the "availabilities" a_{ik} , which each one reflects the accumulated evidence for how appropriate it would be for point i to choose point k as exemplar. Note that for both r_{ik} and a_{ik} point i send them to point k . The messages exchange is repeated until convergence or for a given number of iterations. Then the exemplar of each data point k is point i that maximizes $a_{ik} + r_{ik}$. The process is illustrated in Figure 2.

The algorithm takes as input measures of similarities s_{ik} between data points and the "preferences" p_k which are a real number for each data point. Data points with larger values are more likely to be chosen as exemplars. We have initialized p_k with the median of the input similarities for each data point. We also set the dumping factor λ to 0.5 as the authors propose.

Algorithm 1 Affinity Propagation

Require: $\forall_{i,k} a_{ik}^{-1} = 0, s_{ik} = p_k, \forall_{i,k,i \neq k} s_{ik} = \|x_i - x_k\|^2, \lambda = 0.5, t = 0$
1: **while** !skip **do**
2: $r_{ik}^t \leftarrow (1 - \lambda) \left(s_{ik} - \max_{k', s.t. k' \neq k} \{a_{ik'}^{t-1} + s_{ik'}\} \right) + \lambda r_{ik}^{t-1}$
3: $a_{ik}^t \leftarrow (1 - \lambda) \left(\min \{0, r_{kk}^{t-1} + \sum_{i' s.t. i' \in \{i,k\}} \max \{0, r_{i'k}^{t-1}\} \} \right) + \lambda a_{ik}^{t-1}$
4: $a_{kk}^t \leftarrow (1 - \lambda) \left(\sum_{i' s.t. i' \neq k} \max \{0, r_{i'k}^{t-1}\} \right) + \lambda a_{kk}^{t-1}$
5: $t \leftarrow t + 1$
6: **end while**

4.2. Optimization

Optimization process aims to find an optimal solution to each subproblem. In our problem, complete methods that find the optimal solution are not applicable due to the dimensionality of the problem. Despite we divide the initial ILA problem during the clustering step, some clusters are very big and complete methods are unfeasible. Thus we need to use approximate methods that find a near optimal solution.

The heuristic technique we have chosen is Simulated Annealing (SA) (Russell & Norvig, 2010) according to the study presented in (Arostegui, 2006; Torrent-Fontbona, 2012). It consists in given an initial solution iteratively improves it by selecting another neighbor solution changing the services of some facilities and comparing them according a fitness function. Then it usually moves towards the best solution and repeats the process until convergence or for a given amount of time. The algorithm has the particularity of allowing some "bad moves" that consist in going (sometimes) towards the worst solution to avoid getting stacked on local optimums or flat regions. Its implementation is showed in Algorithm 2, where s is a string of length N_b which in each slot contains the service each facility offers (solution); the fitness function is the target function in section 3, equation (2); τ is a parameter used to adjust the facilities probability to change the offered service. We decided to adjust τ depending on the density of customers (steps 1-5 of Algorithm 2) and depending on the algorithm phase (steps 9-11 of Algorithm 2). Thus, if the ratio $\frac{N_c}{N_b}$ is low, τ is reduced to decrease the probability to change a service because if there are not a lot of customers, facilities occupancy would be lower. Moreover, τ is reduced in the final iterations in order to increase the convergence of the algorithm.

The main inconvenient of using SA is the lack of a metric system in the solution space or a distance function to select neighbor solutions from the current one. Without a coordinate system or a distance function is impossible to know whether two solution are neighbors. Thus, we implemented a new neighborhood function, Algorithm 3, where from a given solution it creates another one assuming it as a neighbor. Such function changes the service of each facility using a certain probability given by an exponential function which depends on the occupancy of the facility o_i and the parameter τ . Using this neighborhood function, facilities with low occupancy have more chances to change their service than facilities with high occupancy. In other words, we consider that the distance between two solutions is given by the number of facilities that offers different services

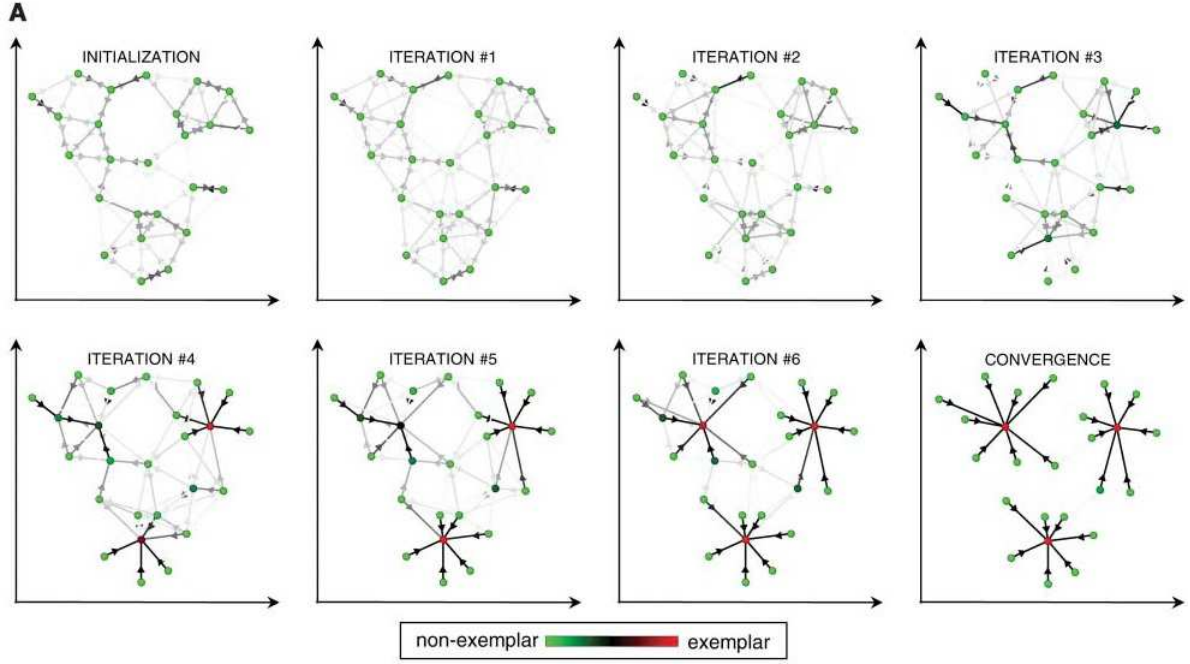


Figure 2: Illustration of the affinity propagation clustering process.

Algorithm 2 Simulated Annealing

Require: $T_a \leftarrow 0.0001$, $T \leftarrow 1$, $\delta T \leftarrow 0.99$, $E_{best} \leftarrow -\infty$, N_c , N_b , $Th_{dens} \leftarrow 15$, $k \leftarrow 10$

- 1: **if** $\frac{N_c}{N_b} < Th_{dens}$ **then**
- 2: $\tau \leftarrow 0.04$
- 3: **else**
- 4: $\tau \leftarrow 0.1$
- 5: **end if**
- 6: Select an initial random solution s
- 7: $E = fitness(s)$
- 8: **while** $T_a < T$ **do**
- 9: **if** $T < kT_a$ **then**
- 10: $\tau \leftarrow 0.02$
- 11: **end if**
- 12: $s' \leftarrow neighbor(s, \tau)$
- 13: $E' \leftarrow fitness(s')$
- 14: **if** $E - E' < 0$ **then**
- 15: $E \leftarrow E'$; $s \leftarrow s'$
- 16: **if** $E_{best} < E$ **then**
- 17: $E_{best} \leftarrow E$; $s_{best} \leftarrow s$
- 18: **end if**
- 19: **else**
- 20: get a random number x between 0 and 1
- 21: **if** $x \leq e^{-\frac{E-E'}{T}}$ **then**
- 22: $E = E'$; $s \leftarrow s'$
- 23: **end if**
- 24: **end if**
- 25: $T \leftarrow \delta T \cdot T$
- 26: **end while**

Algorithm 3 Neighborhood function

Require: s , N_b , N_m , τ

- 1: **for** $i = 0 \rightarrow N_m$ **do**
- 2: get a random number x between 0 and 1
- 3: **if** $x \leq e^{-\frac{\sigma_i}{\tau}}$ **then**
- 4: get an integer random number m between 1 and N_m and $m \neq s_i$
- 5: $s'_i \leftarrow m$
- 6: **else**
- 7: $s'_i \leftarrow s_i$
- 8: **end if**
- 9: **end for**
- 10: **return** s'

5. Experimental Set up

We have applied our method in the barmen decision problem explained in section 1. For the experiments we have used five datasets of bars from Catalonia. The bars correspond to real business taken from *Paginas Amarillas Catalunya*.

- Dataset 1: 373 bars
- Dataset 2: 458 bars
- Dataset 3: 1882 bars

- Dataset 4: 1925 bars
- Dataset 5: 2502 bars

The demand has been simulated generating a random number of customers between 0 and $\frac{2C_i}{3}$ around each bar (where C_i is the bar capacity) and randomly distributed according a bi-dimensional Gaussian distribution centered on bars location. Then each customer decided its wished match according the estimated audience percentage of each match. There were three simultaneous (available) matches with random audience percentages but one of them gathered most of the audience.

To show the benefits of our methodology (Affinity+SA), we have defined our experimental scenario in which no clustering (SA-alone) has been performed. Since no exact optimization method can be applied but an heuristic one, the SA-alone method is expected to perform worse (the search space is big). Due to time constraints we have not used large very big datasets with more than 3000 facilities.

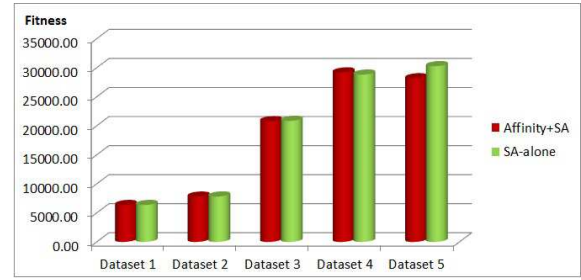
To analyze the benefits of a global strategy (with a fairness mechanism) versus an individual strategy while deciding the optimum match for each bar we have simulated 10 different demands (representing 10 different days) generating around each bar, a random number of customers between 0 and N_{max} where N_{max} is a random number between $1.5C_i$ and $2C_i$. Customers have been distributed around the bars according a bi-dimensional Gaussian distribution like before. Then each customer decided its wished match according 10 different lists of matches (one per day) where matches in each list have assigned a probability of being chosen according the estimated audience of each match.

Experiments have been run on a Intel® Core™ i5 CPU@2.80GHz and 8.00 GB of RAM and Windows 7.

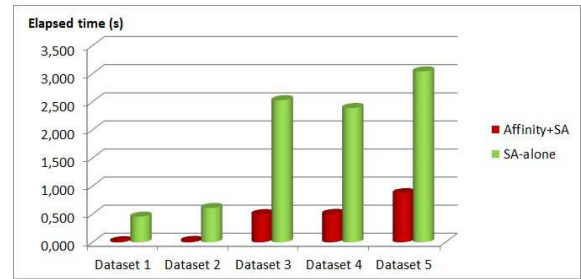
To analyze the achieved solution by the methods we have focused on the fitness of the solutions (the result of the target function, equation (2)), the elapsed time by each method, the occupancy of the bars and the number of allocated customers as a measure of customer satisfaction.

6. Results and Discussion

Figure 3 compares the quality of the solutions and the elapsed time to find them using Affinity+SA or SA-alone. Thus, it confirms that the use of affinity propagation to divide the initial problem does not decrease the quality of the found solutions but, it reduces significantly the elapsed time by SA-alone to find the solution because it reduces the complexity of the allocation process. The reduction has been 95.92% for the



(a) Fitness



(b) Elapsed time

Figure 3: Fitness of the found solutions and the elapsed time by SA to find them.

smallest dataset to 71.25% for the biggest dataset. Figure 4 also shows us that clustering the data space does not reduce the number of allocated customers and it also does not increase the percentage of bars with low occupancy, specially when the problem is big enough. Therefore, clustering the data space helps to solve very big ILA problems keeping the quality of the solution found by another search algorithm like SA.

Despite the number of almost unoccupied bars is low, they can be avoided using a fairness mechanism like the here presented. Fairness included in equation (7) guarantees that if a bar has a low occupancy one day, on next day the system will tend to maximize its occupancy since it has a higher weight. To test our fairness approach the experiments has been repeated 10 times using two algorithms: (i) Algorithm 2 with equation (7) as target function (global strategy) and (ii) an individual strategy based algorithm that consists in selecting, for each bar, the match that maximizes the number of customers, weighting them by $1/1+d_{ij}^2$ without considering the other bars.

We have used 2 different scenarios both with 3 simultaneous matches but with different distribution demand: in the first scenario all matches gather similar audience percentages while in the second scenario one of the matches gather most of the demand. It is expected that for both scenarios the individual strategy presents

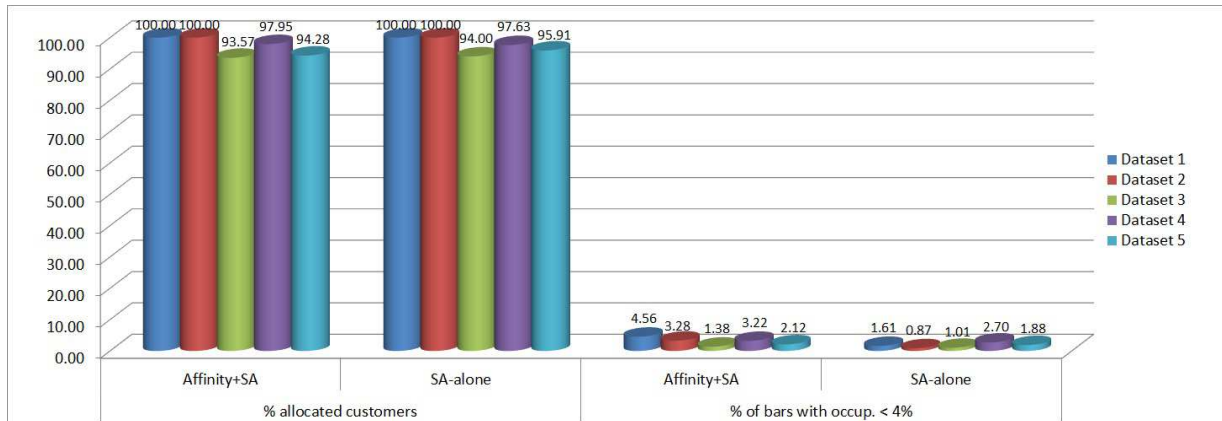


Figure 4: Percentage of allocated customers and bars with occupancy lower than 4%.

higher differences between bar occupancies than the global strategy (SA) due to the fairness mechanism but also we expect a higher mean occupation for the global strategy than for the individual strategy. As we expected, Figure 5 shows how the global strategy achieves higher occupancy percentages for most of the bars while reduces differences between bars in terms of occupancy, o_i . In the individual strategy we can see that for equivalent demand for all matches, some bars remain almost unoccupied because sometimes the best-placed bars attract most of the demand of the region leaving other bars almost empty. But, with our method, all facilities would increase its revenue due to attendance of a higher number of customers as shown by the mean values of Figure 5 (in mean values the occupancy has increased from 48.31% to 82.00% in case (a) and from 68.48% to 80.54% in case (b)). Moreover, the number of satisfied customers has increased from 57.91% to 98.30% in case (a) and from 81.75% to 96.15%.

Regarding the use of other clustering and optimization techniques, see (Torrent-Fontbona, 2012) for an analysis of them which conclude that affinity propagation plus SA are the best combination.

7. Conclusions and Future Work

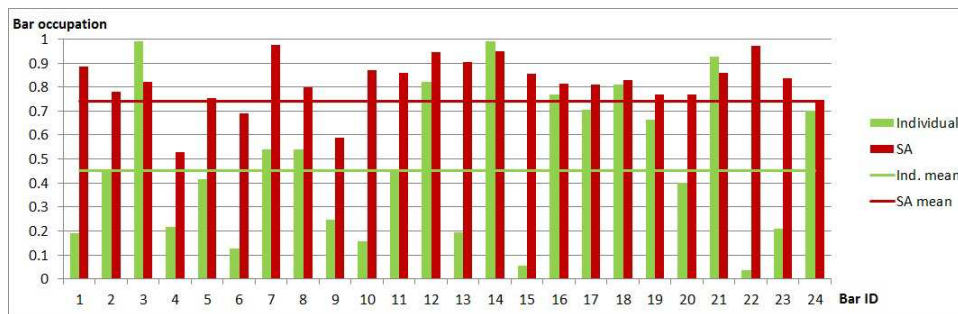
In this paper we tackle the ILA optimization problem which consists in deciding the services each facility should offer to a given demand, and allocate the customers accordingly. Particularly, we revise the problem for a globalization scenario which is posing new challenges to optimization techniques regarding problem dimensionality. For that purpose, first we provide two problem models and then a method to solve them. The first model follows the classical ILA approach, while the

second includes a fair mechanism for repeated problem solving episodes, such that the outcome of one solution feeds the input data of the next problem formulation.

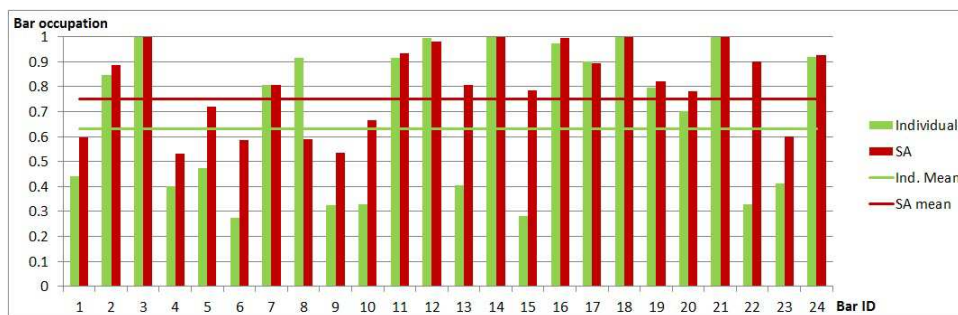
The presented method to solve the problem consists in dividing a big combinatorial problem into smaller subproblems using the clustering technique called affinity propagation and solving each subproblem separately using simulated annealing.

We have tested our method in the barmen decision problem consisting in deciding the best match each bar should offer. We have concluded that our method performs much better than using just a search algorithm (without clustering the data space) in terms of elapsed time while it keeps the quality of the solution. From the bars point of view, we have checked that the mean occupancy of each bar is generally enhanced when using our system instead of deciding the best match for each bar without considering the others.

Although the results are satisfactory there is still room for further analysis as possible inputs. First, we have used the Euclidean distance to measure the bar-customer distance, however, it would be interesting to use the real distance or even the temporal distance to take into account different means of transport. Second, the presented method assumes that customer positions are known just before the beginning of the match. Thus, it could be interesting to include in the system a customer position estimator. Finally, the complexity of our problem is mostly given by the number of facilities and we have used clustering analysis to reduce this complexity. However, the number of customers also increases the complexity of the problem, specifically over the allocation process. To reduce the number of customers, demand aggregation could be used and it would be also interesting to study in a further how this approach af-



(a) 3 matches with similar audience percentage



(b) 3 matches and one of them has between 60% and 80% of the audience

Figure 5: Mean occupancy of 24 different bars along 10 different days using equation (7) as target

fects our methodology.

References

- Aboolian, R., Berman, O., & Drezner, Z. (2008). Location and allocation of service units on a congested network. *IIE Transactions*, *40*, 422–433. Doi: 10.1080/07408170701411385; M3: doi: 10.1080/07408170701411385; 19.
- Aboolian, R., Sun, Y., & Koehler, G. J. (2009). A location allocation problem for a web services provider in a competitive market. *European Journal of Operational Research*, *194*, 64–77.
- Arostegui, M. A. (2006). An empirical comparison of tabu search, simulated annealing, and genetic algorithms for facilities location problems. *International Journal of Production Economics*, *103*, 742–754.
- Brimberg, J., Hansen, P., & Mladenovic, N. (2006). Decomposition strategies for large scale continuous location-allocation problem. *IMA Journal of Management Mathematics*, *17*, 307–316.
- Cooper, L. (1963). Location-allocation. *Operations Research*, (pp. 331–343).
- Frey, B. J., & Dueck, D. (2007). Clustering by passing messages between data points. *Science*, *315*, 972–976.
- GU, W., & Wang, X. (2010). Studies on the performance of a heuristic algorithm for static and transportation facility location allocation problem. In *2nd International Conference on High Performance Computing* (pp. 27–37).
- Melo, M. T., Nickel, S., & da Gama, F. S. (2009). Facility location and supply chain management – a review. *European Journal of Operational Research*, *196*, 401–412.
- Pasandideh, S. H. R., & Niaki, S. T. A. (2010). Genetic application in a facility location problem with random demand within queuing framework. *Journal of Intelligence Manufacturing*, *23*, 651–659.

- Russell, S., & Norvig, P. (2010). *Artificial intelligence: a modern approach*. (3rd ed.). New Jersey: Pearson.
- Torrent-Fontbona, F. (2012). Decision support methods for global optimization. MSc. Thesis, Girona: University of Girona.