



Optimization of real-world outdoor campaign allocations

Fatmanur AKDOĞAN¹, Doğan ALTAN^{2,*}, Ercan PEKER¹, Altuğ ÜSTÜN¹,
Sanem SARIEL²

¹CPM Software, İstanbul, Turkey

²Artificial Intelligence and Robotics Laboratory, Department of Computer Engineering,
İstanbul Technical University, İstanbul, Turkey

Received: 16.07.2019

Accepted/Published Online: 06.02.2020

Final Version: 08.05.2020

Abstract: In this paper, we investigate the outdoor campaign allocation problem (OCAP), which asks for the distribution of campaign items to billboards considering a number of constraints. In particular, for a metropolitan city with a large number of billboards, the problem becomes challenging. We propose a genetic algorithm-based method to allocate campaign items effectively, and we compare our results with those of nonlinear integer programming and greedy approaches. Real-world data sets are collected with the given constraints of the price class ratios of billboards located in İstanbul and the budgets of the given campaigns. The methods are evaluated in terms of the efficiency of the constructed plans and the construction time of the planning. The results reveal that the genetic algorithm-based approach gives close to optimal results in the shortest scheduling time for the OCAP, and it scales linearly with the increasing data sizes.

Key words: Out-of-home advertisement, billboard scheduling, campaign allocation, genetic algorithms, optimization

1. Introduction

Out-of-home advertising, with its various advantages over the other advertisement mediums, for which ignorance, zapping, or clicking away is more likely, is one of the oldest communication tools to reach business and consumer markets. Printed or digital outdoor advertisement signs are commonly used by organizations to communicate their campaigns outdoors. These could be displayed on billboards, street furniture, mobile vehicles, etc. [1]. In the present study, we particularly focus on printed billboard advertising and the allocation of advertisement campaign posters to billboards. The methods that we present are also applicable to all forms of outdoor advertising formats.

Billboard advertising is one of the oldest methods of out-of-home media. It appears in various forms such as citylight posters, megalights, electric poles, shopping mall kiosks, and bus stop posters. This form is mostly seen in city centers. For instance, bus stop posters are displayed at bus stops that are visible to people waiting for a bus. Billboards are usually located in heavy traffic areas such as high-visible highways, main arteries and intersections. Billboard ads with high visibility have a significant impact on traffic vehicles.

Repetition in advertising is an important tool in out-of-home advertising [2]; that is, the higher number of campaign items to display, the higher number of customers you reach. On the other hand, display locations of campaigns are more crucial than how frequently they are displayed. Hence, it is of great importance to allocate advertisement items of a campaign by considering the geographic concentration, the demographic profiles of the

*Correspondence: daltan@itu.edu.tr

audience, or the specific locations close to business centers. In the present study, we aim to develop appropriate campaign item assignment methods for out-of-home billboards.

We call this investigated problem the outdoor campaign allocation problem (OCAP). An illustration of the problem and assignments are given in Figure 1. Campaign items need to be distributed to the billboards considering the distribution rates. Billboards may be classified under different categories considering their locations. Moreover, each billboard has location information (e.g., continent information for İstanbul). The OCAP can also be represented as an online matching problem with weighted edges and capacity limits. The costs and power of reaching the audience with that billboard can be represented by weighting the edges on the graph. This problem differs from the media planning problem [3, 4], where a decision is given on which medium or type of advertising to use by small companies with limited budgets and little technical knowledge.

The campaign item requests in the OCAP are collected through the advertising agencies of customer companies. These are distributed to the advertising channels according to their budget, and an allocation is requested from all advertising channels. Although the display periods of advertisement items can change depending on the product type (billboard, megalight, etc.), the campaigns are usually planned as monthly and weekly.

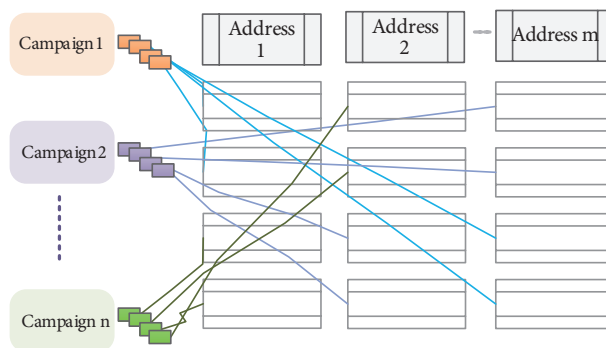


Figure 1. Outdoor campaign allocation problem.

In the manual allocation process by an operator, the best combination of the addresses is selected for the customer to ensure reaching the right audience of the requested campaign. The following considerations are taken into account: How many people can be reached by distributing items to different product types? Which media platforms and product types should be allocated? Does the distribution of the campaign items match the profit from the assigned addresses?

Considering this information, an operator has to allocate items for at least 30 customers per day on average. Moreover, for each campaign request sent by the sales representative, the agency may request a revision to an existing allocation on average 30 times a day due to location, product type, period, and changes occurring.

In the meantime, 80% of the incoming campaign requests are planned as optional, while only 20% of them are planned as real sales most of the time. Indeed, taking into account that even the status of sales may change, it may be necessary to design a continuous revision process until all campaign requests are finalized. This entails a great workload for the operators.

In the present study, methods for performing automated campaign-billboard assignment approaches are investigated, and the proposed methods are applied and analyzed on real-world data sets. The OCAP may

be addressed by linear integer programming frameworks to achieve the optimal results. Greedy algorithms [5] that do not guarantee optimal solutions but give acceptable solutions, or heuristic methods can also be used to solve the problem. We propose a genetic algorithm (GA)-based method that can provide close to optimal solutions within the shortest allocation time. We also provide a comparison of our method with manual planning, nonlinear integer programming, and greedy approaches. Detailed analyses of these methods are presented for various real-world data sets. Our main contribution relies on solving the OCAP, which has a real-world domain with a GA-based methodology. The data set that we use consists of real customer requests from the year 2018 in İstanbul, and the proposed method can generate solutions for these requests within the shortest allocation time.

This paper is organized as follows. First, the problem is introduced and the literature on this problem is presented. Then the problem is formulated in a formal manner. This section is followed by the presentation of the methods to handle the corresponding problem: manual planning, integer programming (IP) formulation, greedy algorithm-based solution, and GA-based solution. Then an experimental evaluation is presented on the real-world data. Last, the paper is concluded with possible future directions.

2. Related work

Media allocations that we are interested in herein are usually represented as matching problems. The stable marriage problem [6], student-dormitory assignment problem [7, 8], kidney transfer assignment problem [9], worker-staff assignment problem [10, 11], hospitals/residents problem [12], and college admissions problem [6] are examples of well-known matching problems relevant to our studies for the OCAP.

An IP model may also be used to handle the problem with natural heuristics for a multiperiod problem. Adany et al. [13] present a solution for a media allocation problem and compare the results with those of a heuristic method and a deterministic method. In another work, a display ad allocation problem is handled by using a linear programming (LP) formulation, where online bidding algorithms are proposed [14]. Measurement of reaching the audience or audience frequency is also a key problem of outdoor advertising and marketing, and it is handled in Lichtenthal et al. [1]. Trajectories of the billboards and the campaigns to be assigned to them may also be taken into account in order to maximize the influence of an advert [15] by comparing a greedy algorithm, a partition-based framework, and a LazyProbe method. In another work, data collected from the mobile phones of users are used to construct an optimal assignment for the digital advert-billboard assignment problem [16].

The OCAP is formulated as an optimization problem based on the needs of the software company CPM. The problem most similar to this one is the media planning problem, where the objective is to decide on the best advertisement media for a company. This is a complex problem considering its parameters such as media revenue, limited area inventory, demographic characteristics, different target groups, customer preferences, and competition [5, 17]. Although there are many media channels to reach the customers, similar problems are faced regarding similar parameters in every channel. TV advertisements are common media tools, and Cosgun et al. [18] developed a mathematical model for the problem called an advertisement reservation problem by using a mixed integer LP approach to solve how to reserve the television advertisements by customers. Similarly, Zhang [19] proposes a mathematical model to solve the television advertisement allocation problem using two-step hierarchical approaches.

In particular, online and internet advertisement allocation is a challenging problem [5, 14, 17, 20–23]. The advertisements are allocated based on auctions or contracts and supply/demand constraints are taken into

account to determine the best solutions. Our inspiration for the OCAP representation comes from the media planning domain. The media planning problem has been represented as a weapon target assignment (WTA) problem previously, and the solutions are generated with various optimization methods [24]. This problem aims to find a feasible assignment between weapons and targets considering maximization of the expected damage. The assignment is also considered as an NP-complete problem. This problem can be solved using an IP approach. However, as the problem size grows, it can be handled by a GA-based solution [25]. Although the OCAP differs from this problem, we are inspired by the presented methodology for solving the media planning problem in these works. Our formulation and solution target solving the OCAP but can be extended to solve other optimization problems as well.

3. Outdoor campaign allocation problem

The main objective of the present work is to allocate advertisement campaign items to the existing sets of billboards distributed and located in different parts of the city. The following definitions are needed to present our formulation to solve the OCAP.

- **Billboard:** A billboard is an outdoor structure where posters are displayed. Each billboard may have a different number of faces to display posters in a time-sharing manner.
- **Address:** An address of billboards represents an area where a group of billboards are located.
- **Campaign:** A campaign corresponds to an advertisement decision for a company for promotion and marketing.
- **Campaign poster:** A campaign poster is an illustration that is displayed on a billboard. Campaign posters are the tools to promote a campaign.

Based on these definitions, the OCAP can be formulated as the following matching (M) problem. Given a graph $G(V, E)$ of nodes, which is divided into two sides with one side involving campaigns from each company i ($c_i \in C$) and the other side involving billboards ($b_j^l \in B$), the matching must satisfy a certain criteria set. Each campaign c_i calls for a number (v_i) of posters (each symbolized as p_k^i) are allocated to the billboards for advertisement with the same unit price u_{c_i} .

Billboards are located in groups at certain addresses ($a_l \in A$, $b_j^l \in B$). Addresses may have different visibility scores (r_{t_m}), depending on their classes ($t_m \in T$). In our particular case study for CPM, $\|T\| = 3$. The class of an address and its score depend on its location (the higher the value is the better class of address). These scores are mostly determined by an impact measure called visibility adjusted contact (VAC) value for outdoor advertising¹. This value is an indicator for the coverage or reach of a campaign for the corresponding address, and it is dependent on the locations, orientations, and other physical characteristics of displays (e.g., under lightning), the volume of traffic (targeting either cars or pedestrians) at that particular location, and the visibility of other surrounding items.

The matching problem asks for finding an allocation of campaign posters to billboards. Each poster p_k^i is matched with a different billboard b_j^l . However, multiple copies of posters can be displayed in the same address

¹Cafas usage manual [online]. Website <http://www.outdoorimpact.lv/wp-content/uploads/2013/09/Norway-Cafas-usage-manual1.pdf> [accessed 24-May-2018]

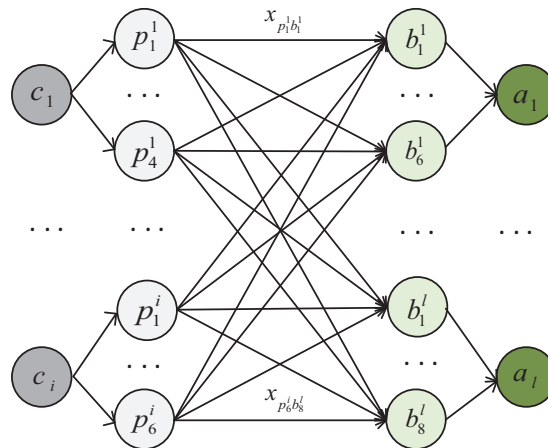


Figure 2. Graph representation of the outdoor campaign allocation problem.

a_l (on a different billboard). The final matching eventually associates a binary value $x_{p_k^i b_j^l}$ based on a set of constraints considering an objective function. This variable represents whether the k^{th} poster of campaign i (p_k^i) is assigned to the j^{th} billboard at address l (b_j^l). An illustration of an example graph is given in Figure 2. In this figure, the leftmost nodes are the campaigns, and they are associated with the given number of posters, where the rightmost nodes are the addresses for the given number of billboards. Note that the values for the number of posters and billboards are chosen from small numbers randomly for simplicity.

Considering the defined variables, there are a number of constraints that need to be satisfied during the allocation process. First, each campaign has an exact copy of posters, and all of them should be distributed to the billboards in the city. Since each address is limited to display only a given number of posters, this capacity cannot be exceeded.

According to the needs of the companies, some soft constraints may exist over the distributions of the posters to the addresses. One common constraint involves obeying the given distribution ratios for address classes (d_{t_m}). Let $v_i^{t_m}$ denote the total number of posters of campaign c_i assigned to the billboards of class t_m . In our particular case study, the campaign posters are distributed to addresses with even number of copies (as the multiples of 2) to increase their visibility with repetition. Given these constraints, the main objective can be summarized as maximizing the overall profit of the company that achieves the allocation considering the unit prices u_{c_i} as follows:

$$maximize(\sum_i (u_{c_i} * \frac{\sum_m (r_{t_m} * v_i^{t_m})}{v_i})) \tag{1}$$

4. Solution methods for the OCAP

In this section, three methods for solving the OCAP are presented. These methods comprise: (i) a nonlinear IP-based optimization method, (ii) a greedy approach, and (iii) a GA-based optimization method.

4.1. Nonlinear IP formulation

The OCAP can be solved by using a nonlinear IP formulation. The main aim is to optimize the allocation of the campaign posters to the billboards as presented in Equation 2. We use a generalized reduced gradient (GRG)

algorithm as a nonlinear IP solver. Depending on the distribution of the campaign posters and considering the billboard classes that they are assigned to, the objective function can be constructed as follows:

$$\text{maximize}(\sum_i (u_{c_i} * \frac{\sum_m (r_{t_m} * v_i^{t_m})}{v_i})) \quad (2)$$

In this equation, the fraction inside the summation returns a weight between 0 and 1 depending on the visibility of the location. A high value shows that the majority of the campaigns are assigned to the billboards with a high class value (t_m). Note that the r_{t_m} value is predefined for each billboard type.

In order to formulate the presented problem as a nonlinear IP formulation, the given constraints are encoded. Each address has a capacity in terms of the campaign posters to be displayed. This capacity constraint can be formalized with Equation 3 for each billboard b_j^l at address a_l .

$$\forall l (\sum_i \sum_k \sum_j x_{p_k^i b_j^l} \leq \text{capacity}(a_l)) \quad (3)$$

All posters of a campaign must be allocated considering the number of copies demanded. Equation 4 represents this constraint for each campaign c_i . Note that multiple copies of campaign posters may be assigned to multiple addresses.

$$\forall i (\sum_k \sum_l \sum_j x_{p_k^i b_j^l} = v_i) \quad (4)$$

Each campaign poster must be distributed to a billboard at an address exactly once (Equation 5), and each distinct billboard face at an address must contain exactly one poster (Equation 6).

$$\forall i \forall k (\sum_l \sum_j x_{p_k^i b_j^l} = 1) \quad (5)$$

$$\forall l \forall j (\sum_i \sum_k x_{p_k^i b_j^l} = 1) \quad (6)$$

While assigning campaign posters to the billboards, class distribution rates for each class type (d_{t_m}) must be satisfied for each campaign. Reaching the exact rate is not possible, but at least an approximation would be possible. Equation 7 formalizes this condition.

$$\forall i \forall m (\text{minimize}(|d_{t_m} - \frac{v_i^{t_m}}{v_i}|)) \quad (7)$$

4.2. Greedy algorithm-based solution

Greedy approaches constitute intuitive solutions for solving optimization problems. A solution with a greedy algorithm is computed in an iterative manner by repeating the same greedy methodology. The following issues are taken into consideration in the design of the greedy algorithm for the OCAP. First, it is necessary to allocate campaign posters with higher unit prices to higher class addresses. Second, at the same time, to spread out the campaign promotion to a great extent, the posters must be distributed over a wide area, particularly to the best

class addresses. This is the case for all campaigns. Moreover, the distribution of posters to address classes must satisfy the given target distribution rates. Although a greedy approach provides a low cost solution, it does not always guarantee the optimal solution. The proposed greedy approach applies the following main steps:

1. *Sorting the campaigns:* In order to give priority to the campaigns that have higher unit prices, the campaigns are sorted based on their unit prices in a descending manner.
2. *Sorting the addresses:* Addresses are sorted based on their classes in descending order.
3. *Calculating distribution rates:* In order to maintain the distribution rate of each class for each campaign, a list is constructed in order to include amounts to fulfill these rates.
4. *Assigning campaign posters to the billboards:* Using the constructed address list in the previous step, each campaign is assigned to a billboard at each execution step of the algorithm.

4.3. GA-based solution

GAs are evolutionary optimization methods that iterate over a number of possible candidate solutions to find a result without any guarantee of an optimal solutions. A population of individuals (chromosomes) is maintained to represent a number of possible solutions to the problem. Individuals are made of genes, each of which corresponds to part of the solution. Nature inspired genetic operators, namely *selection*, *crossover*, and *mutation*, are used to iterate the search process over these individuals. Selection is applied to select individuals that undergo crossover and mutation. Crossover simply exchanges a block of genes between two individuals. Mutation is applied to randomly change a gene's value, to mutate the representation of the individual.

The proposed GA-based algorithm for solving the OCAP represents each individual to correspond to a campaign. Since a campaign is supposed to be displayed on many billboards, there is a gene block for each campaign. For example, if a campaign is supposed to be displayed 10 times, then 10 sequential genes are reserved for this particular campaign. Therefore, each billboard (including its faces) has a distinct ID that represents it. Example individual solutions are illustrated in Figure 3. We need to allocate 3 campaigns with different numbers of face requirements (for displaying corresponding campaign posters), 2, 4, and 4, respectively. Assume that there are 2 addresses with different number of billboards, 4 and 6 in this example. A sample solution includes 10 genes in total. The first two genes of the individual are reserved for the first campaign, the next four genes are for the second campaign, and so on. Note that the gene clusters for each campaign are assigned to a different color in the figure. Each integer in a gene denotes the billboard ID of the corresponding address. For example, for the first individual in the figure, the first campaign's first poster is allocated to the first address, to appear on the second and fourth billboards. The second campaign's first poster is allocated to appear on the sixth billboard (its ID is: 10) in the second address.

After the initial population is constructed randomly, at each iteration of the GA, parent selection, crossover, and mutation operators are applied sequentially. Partially mapped crossover (PMX) is used as a crossover method in order to secure the feasibility of an individual since random exchanges may lead to infeasible solutions (e.g., a particular copy of a campaign poster can only be published on a single billboard and a billboard can only contain a single poster.). Figure 3 illustrates the PMX operation on a randomly selected individual pair. In this crossover method, two points on the individuals are randomly selected to exchange gene numbers. In the illustration, genes from 5th to 7th are selected for the crossover operation. Since every campaign should be assigned to a unique billboard ID, a repair procedure may be needed to fix the infeasibility

of an individual. In this example, genes that are labeled with 3, 8 and 1, 2 cause violations and need to be repaired. These fixed genes are marked with an asterisk. These genes need to be swapped since the individuals become infeasible candidates after the partial swap due to having duplicates of the same campaign. These genes are determined by using a map considering the swapped genes during the crossover operation. Swap mutation is used as a mutation operator where two genes are randomly swapped. In this mutation procedure, two genes are selected at random and they are interchanged. Since the swap operation only changes the order, this mutation procedure does not violate feasibility. An example mutation operation is illustrated in Figure 4, where the genes marked with asterisks denote the interchanged ones. Roulette wheel selection is applied to select the new generation from the current population.

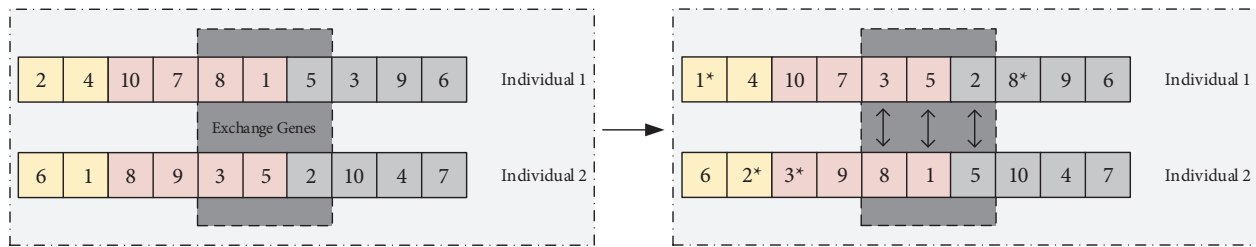


Figure 3. Crossover operation on two individuals.



Figure 4. Mutation operation on an individual.

$$\text{maximize} \left(\sum_i \left((u_{c_i} * \frac{\sum_m (r_{t_m} * v_i^{t_m})}{v_i}) - (f_{pen} * \frac{\sum_m |d_{t_m} - \frac{v_i^{t_m}}{v_i}|}{|T|}) \right) \right) \tag{8}$$

This overall process is summarized in Algorithm 1. The GA starts with generating the initial population randomly. Then, for each generation, two pairs of individuals are selected from the existing generation to undergo crossover. After the crossover is applied to the pairs, each individual is considered for mutation. Fitness calculation of an individual is presented in Equation 8. Note that f_{pen} is the penalty coefficient. The first part corresponds to the summation of the multiplication of the unit prices with the class coefficients for each campaign. The subtraction part is the penalty that considers the difference in the actual distribution rate and the predefined distribution rates for each campaign. Based on this computation, the fittest individual is selected to be transferred to the next generation (without undergoing the selection procedure), where the other individuals go through to the next generation using binary roulette wheel selection by considering the fitnesses of the candidates, that is, each individual has a survival chance proportional to its fitness. The remaining spots in the new generation are filled with randomly constructed individuals.

Algorithm 1: GA for the OCAP

Data: Population size p , Generation limit g , Crossover rate c , Mutation rate m
Result: A feasible assignment s

```

1  $P = \text{generateInitialPopulation}();$ 
2 while  $epoch \leq g$  do
3   for  $i \in P$  do
4      $i_1, i_2 = \text{selectIndividual}(P);$ 
5      $\text{applyCrossOver}(i_1, i_2, c);$ 
6      $\text{applyMutation}(i_1, i_2, m);$ 
7   end
8    $fittest = \text{findFittest}(P);$ 
9    $P = \text{selectNewGeneration}();$ 
10   $\text{addFittest}(P, fittest);$ 
11 end
12  $s = \text{findFittest}(P);$ 
13 return  $s;$ 

```

5. Experiments

In this section, we present the real-world data set used in the present study and the results of the experiments.

5.1. Experimental setup

The real-world data set is directly taken from real allocation instances provided by CPM. The company has a wide product range that provides services to almost every province in Turkey. The product range includes products like city light posters (CLPs), billboards, electric poles, tower boards, mega boards, megalights, metro boards, poster boards, roller towers, and bus stations. The instances are collected from real campaign allocations in 2018. The company has a total of nearly 40,000 ad units. In our study, we focus on the allocation of billboards in İstanbul as a case study. There are 913 billboards in total and 4690 advertising face capacities. Of this volume, only 3044 of them are used for pilot applications.

The largest data set, with 3044 campaign posters, is from 18 different sector types with different numbers of campaigns. The main sectors in this large data set are as follows: real estate, fair, retail, durable consumption, education, finance, accessories, IT, social, tourism, automotive, textile, food, sports centers/equipment, media, and municipality. To evaluate the scalability of the algorithms, three smaller data set groups (each of which including four randomly constructed data sets) with various sizes are constructed at random from the real data as follows: DS_1 : 5 different addresses and 5 different campaigns (with data set IDs 1-4), DS_2 : 10 different addresses and 10 different campaigns (with data set IDs 5-8), and DS_3 : 20 different addresses and 10 different campaigns (with data set IDs 9-12). Note that the largest data set, which includes the entire data, is denoted with data set ID 13.

Figure 5 presents a statistical analysis of the data used in the experiments. Each subfigure presents the characteristics of the data for each campaign in the related data set in terms of quantity and unit price. Each bar set enumerated with a number on the x axis denotes a distinct campaign in the related data set. The leftmost y axis represents unit price while the other (rightmost) denotes the quantity from the corresponding campaign. Figure 5a illustrates the statistics for the data set that includes 5 addresses and 5 campaigns, Figure 5b shows the one that includes 10 addresses and 10 campaigns, and Figure 5c corresponds to the data set with 20 addresses and 10 campaigns. Note that each campaign has multiple copies of posters to be assigned to the

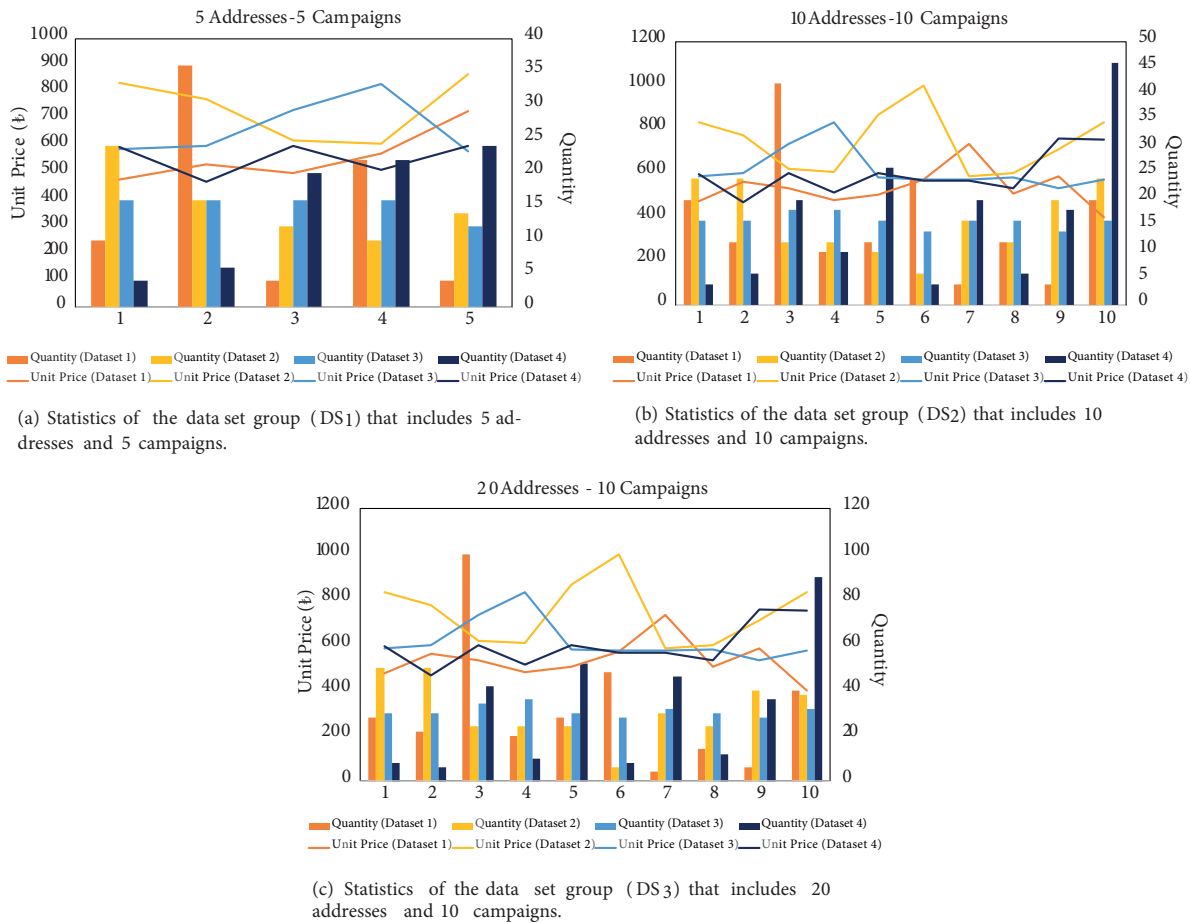


Figure 5. Statistics of the data sets that are used in the present study.

billboards, where each address has several billboards with a number of faces. Therefore, a total of 322 distinct addresses and 41 campaigns exist in our data set. Moreover, each address has a different number of billboards and each campaign has various copies to be published. As seen from the subfigures, each data set has a different scale considering the unit prices and quantities for the campaigns. In the following section, analyses with the proposed methods on these data sets are presented.

5.2. Manual planning process

The motivation behind our work is to automate the allocation phase, which was done manually by operators previously. In the course of these studies, according to the natural process of allocations, we have to pay attention to the demanded numbers and the occupancy rates of the installed products in the field. While operating on the 5-address data set, it is not challenging to distribute the demanded amount of posters homogeneously due to the low number of installed products and clients. Since in the first distribution stage the demanded campaign poster quantities and the occupancy rates of the installed products are not taken into consideration, the constraints may be violated. Consequently, after the distribution of the campaign posters is complete, an improvement is made for resolving these constraints.

In the 10-address data set, while allocating, due to the random order of the address levels and the increasing number of campaigns, the operator needs to focus on combinations of the bookings. One of the

reasons that make the allocations difficult is the following: while trying to pay attention to the number of installed products at the addresses and to the homogeneous distribution quantities, in the meantime it is difficult to allocate the items. Considering these, the allocations for 10 campaigns could be distributed according to the address classes, but the number of installed products at each address and the required quantities for other campaigns could lead to incomplete or excessive planning. Likewise, by the nature of media planning, backtracking may be required for each distributed campaign poster, and even reallocation may be necessary. This situation again prevents both the efficiency of time management and effective allocation. In the 20 address data set, it is necessary to be more careful while allocating since the number of addresses increases. In terms of homogeneous distribution, it is even more challenging to reach the effective distribution of each customer and follow the quantities remaining at the addresses. In some cases, while capturing the distribution rates of the customers, the remaining quantities in the addresses may fail to be noticed and overplanning may occur. Taking into account all of these, allocations of a campaign may result in ineffective time management. As a result, for each data set, it can be challenging to track the slots, to control the homogeneous distributions of the requested quantities for the campaigns, to assign them to the right locations, and also to complete the process by considering the needs of the new campaigns.

5.3. Experimental results

In this subsection, we present experimental results on the presented data sets by the described methods. The GA-based method is run 500 times for each data set with the following parameters: the number of generations as 500, the crossover rate as 0.7, the mutation rate as 0.1, and the penalty coefficient (f_{pen}) as 6000. Table 1 presents the performance analysis of the methods. In each column, data set ID, the number of billboards (addresses) and the number of campaigns that the corresponding data set includes, and the results of the methods are presented. Each row shows the results for the corresponding data set in terms of the quality of the solution (best fitness), the average deviation rate from the billboard class distribution rates, and the time elapsed for the related method to come up with the solution. In addition to the data sets explained in Figure 5, the last data set row presents the results for a larger data set that includes 322 addresses and 41 campaigns to be assigned. The last row represents the average deviation rates from the expected billboard class distribution rates for each method. As can be seen from Table 1, for the small data set group, all of the methods perform similarly. Note that the manual planning method corresponds to the solution where a human operator assigns the campaigns to the billboards. It can also be inferred from the table that the manual planning and IP methods require more time compared to the other methods. As the data set size grows, the GA-based solution results in close to optimal results in terms of best fitness. In some cases, heuristic methods (the GA-based and the greedy methods) seem to find solutions better than the optimal ones found by the IP method. However, note that in these cases there is a small amount of violation (as soft constraints) in the average deviations of the class distributions that are acceptable by the company (less than 15%). It has also been observed that the results of the greedy method have the highest deviation rates. As the numbers of campaigns and billboards grow, the manual planning and IP-based methods cannot even come up with a result. For example, the last data set row indicates the results of the largest data set with 322 billboards and 41 campaigns. By the manual planning and IP-based methods, it is not possible to generate a solution within an acceptable period of time for this data set. On the other hand, the greedy algorithm and the GA-based methods can provide solutions. Overall, the GA-based solution generates acceptable solutions within reasonable time periods and therefore is more efficient than the others.

Table 1. Performance analysis of the methods on the real-world data sets.

# of addresses	# of campaigns	Data set ID	Manual planning		Integer programming		Greedy algorithm		Genetic algorithm	
			Best fitness	Time (s)	Best fitness (dev. rate)	Time (s)	Best fitness (dev. rate)	Time (s)	Best fitness (dev. rate)	Time (s)
5	5	1	1634.30	162.89	1634.30 (6%)	7.06	1641.03 (7%)	2.15	1632.70 (6%)	0.04
		2	1988.77	138.82	1988.77 (3%)	3.60	1975.85 (16%)	2.49	1997.45 (4%)	0.04
		3	1781.24	130.51	1781.24 (3%)	5.73	1805.10 (14%)	2.95	1788.58 (3%)	0.04
		4	1526.93	145.34	1526.93 (5%)	2.98	1498.27 (12%)	2.33	1524.97 (6%)	0.04
10	10	5	2932.20	360.16	2932.20 (5%)	106.42	2928.97 (8%)	7.78	3032.94 (6%)	0.09
		6	3841.80	330.33	3841.80 (4%)	98.00	3918.20 (6%)	7.85	4074.56 (6%)	0.09
		7	3209.06	176.00	3209.06 (4%)	127.25	3213.63 (4%)	9.15	3238.38 (5%)	0.09
		8	3210.57	470.00	3210.57 (6%)	82.29	2966.45 (16%)	7.84	3329.40 (7%)	0.09
20	10	9	2951.55	803.70	2951.55 (4%)	361.50	3049.43 (9%)	14.00	3076.41 (5%)	0.20
		10	3979.43	513.15	3979.43 (3%)	490.92	4124.79 (5%)	14.20	4168.23 (4%)	0.21
		11	3286.67	771.16	3286.67 (3%)	574.65	3285.44 (3%)	15.45	3304.14 (4%)	0.21
		12	3115.52	580.46	3115.52 (4%)	995.92	2970.99 (10%)	13.57	3350.16 (5%)	0.21
322	41	N/A	N/A	N/A	N/A	5055.89 (14%)	355.23	5478.44 (12%)	7.32	
	Avg. dev. rate	4.17%	4.17%	9.54%	5.62%					

Tables 2-4 present an analysis of the parameters of the GA-based solution for each data set considering their sizes. This analysis is based on varying crossover rates and the penalty coefficients f_{pen} . Note that the mutation rate is set to 0.1 during this analysis. In these tables, the key word *Total* represents the total number of campaign posters and billboards (considering the multiple faces of the billboards and copies of the campaign posters). The columns of the table present results for the data sets where each row corresponds to a distinct setting of the GA. According to Table 2, almost all of the crossover settings give similar solutions for both penalty coefficients. For the medium-sized data set group (Table 3), the crossover rate of 0.5 provides better solutions in general for $f_{pen} = 6000$. For the larger data sets (Table 4), the crossover rates of 0.5 and 0.7 are sufficient for finding the best solutions when the penalty coefficient is high. Moreover, for the last and the largest data set, a crossover rate of 0.7 gives the best result. For Tables 2-4, as the penalty coefficient decreases to 100 (in other words, when the class distribution rate constraints are softened), the best fitness values increase because of the allowance of deviations from class distribution rates. For instance, the average deviation from the class distribution rate (d_{tm}) of the solutions for the small-sized data set group increases from 0.05 (when $f_{pen} = 6000$) to 0.21 (when $f_{pen} = 100$), for the medium-sized data sets from 0.06 ($f_{pen} = 6000$) to 0.17 ($f_{pen} = 100$), and for the larger data sets from 0.05 ($f_{pen} = 6000$) to 0.17 ($f_{pen} = 100$).

Table 2. Analysis of the GA-based method for varying parameter settings on the small-sized data sets (DS_1).

		Campaigns	5			
		Addresses	5			
		Total	76			
Penalty	Crossover rate	Data set ID	1	2	3	4
$f_{pen} = 100$	0.5	Duration	0.153	0.151	0.151	0.153
		Best fitness	1970.05	1997.45	1788.58	1726.50
	0.7	Duration	0.176	0.174	0.173	0.1750
		Best fitness	1970.05	1997.45	1788.58	1726.50
	0.9	Duration	0.202	0.196	0.196	0.201
		Best fitness	1970.05	1997.45	1788.58	1726.50
$f_{pen} = 6000$	0.5	Duration	0.020	0.020	0.019	0.021
		Best fitness	1632.70	1997.45	1788.58	1553.60
	0.7	Duration	0.040	0.040	0.040	0.039
		Best fitness	1632.70	1997.45	1788.58	1524.97
	0.9	Duration	0.196	0.198	0.202	0.196
		Best fitness	1632.70	1997.45	1788.58	1524.97

Figure 6 illustrates the evolution of individuals in a population during the GA's generations for the largest data set (a total of 3044 campaigns). In the plot, the x axis represents the number of generations, while the y axis presents the fitness values. The blue line denotes the fitness of the best individual in the current population found so far, while the orange line presents the average fitness of the individuals. As seen from the plot, the

Table 3. Analysis of the GA-based method for varying parameter settings on the medium-sized data sets (DS_2).

		Campaigns	10			
		Addresses	10			
		Total	160			
Penalty	Crossover rate	Data set ID	5	6	7	8
$f_{pen} = 100$	0.5	Duration	0.346	0.344	0.047	0.344
		Best fitness	3453.62	4421.57	3243.07	3767.74
	0.7	Duration	0.474	0.433	0.402	0.425
		Best fitness	3442.52	4402.85	3246.05	3790.12
	0.9	Duration	0.461	0.462	0.461	0.452
		Best fitness	3496.30	4420.69	3245.89	3760.87
$f_{pen} = 6000$	0.5	Duration	0.047	0.047	0.046	0.046
		Best fitness	3023.56	4135.76	3244.86	3337.74
	0.7	Duration	0.094	0.091	0.093	0.093
		Best fitness	3032.94	4074.56	3238.38	3329.40
	0.9	Duration	0.462	0.455	0.458	0.508
		Best fitness	2997.00	4051.30	3235.15	3316.54

solution is generally obtained when the number of generations is approximately 350. Note that the average fitness of the population does not converge to the best fitness value. This is because during the reproduction of the population with binary roulette wheel selection only one individual (the better one) survives. Therefore, the empty spots of the individuals that do not survive are replaced with randomly constructed individuals in the new generation.

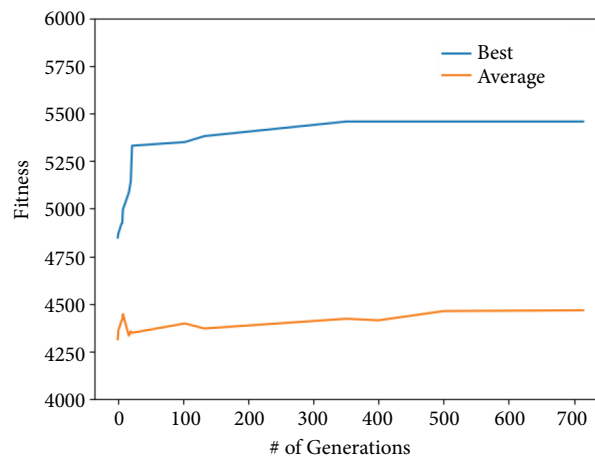
**Figure 6.** Evaluation of the best and average solutions through the GA.

Table 4. Analysis of the GA-based method for varying parameter settings on the big-sized data sets (DS_3) and the complete data set.

		Campaigns	10				41
		Addresses	20				322
		Total	310				3044
Penalty	Crossover rate	Data set ID	9	10	11	12	13
$f_{pen} = 100$	0.5	Duration	0.742	0.749	0.738	0.736	21.905
		Best fitness	3515.95	4423.03	3308.17	3832.26	6019.78
	0.7	Duration	0.886	0.907	0.881	0.889	28.418
		Best fitness	3514.73	4433.16	3312.27	3831.53	5976.06
	0.9	Duration	1.042	1.051	1.033	1.045	35.976
		Best fitness	3524.31	4425.67	3311.40	3816.46	6075.88
$f_{pen} = 6000$	0.5	Duration	0.106	0.105	0.105	0.106	3.556
		Best fitness	3062.92	4159.37	3306.51	3365.55	5378.38
	0.7	Duration	0.205	0.208	0.208	0.210	7.321
		Best fitness	3076.41	4168.23	3304.94	3350.16	5478.44
	0.9	Duration	1.060	1.047	1.030	1.030	35.530
		Best fitness	3063.91	4117.51	3305.64	3305.09	5360.84

6. Conclusion and future work

In this paper, the OCAP is investigated and analyzed. The problem is represented with a GA formulation and tested on various sized data sets that include real-world data gathered from İstanbul. The proposed method is also compared with manual planning-based, IP-based, and greedy algorithm-based solutions. The results indicate that the GA-based method generates close to optimal solutions. Furthermore, when the data size grows, the GA-based solution can handle the problem faster within a few seconds compared to the other methods. Future work includes applying machine learning methods to consider the features of the campaigns to associate with the locations of the billboards while achieving the allocation task.

Acknowledgment

This work was supported by the Scientific and Technological Research Council of Turkey (TÜBİTAK) under Grant #7161058.

References

- [1] Lichtenthal JD, Yadav V, Donthu N. Outdoor advertising for business markets. *Industrial Marketing Management* 2006; 35 (2): 236-247.
- [2] Committee OA. *Essentials of Outdoor Advertising*. New York, NY, USA: Association of National Advertisers, 1958.

- [3] Little JD, Lodish LM. A media planning calculus. *Operations Research* 1969; 17 (1): 1-35.
- [4] Kelley L, Jugenheimer D, Sheehan KB. *Advertising Media Planning*. New York, NY, USA: Taylor and Francis, 2012.
- [5] Mehta A. Online matching and ad allocation. *Foundations and Trends® in Theoretical Computer Science* 2013; 8 (4): 265-368. doi: 10.1561/04000000057
- [6] Gale D, Shapley L. College admissions and the stability of marriage. *American Mathematical Monthly* 1962; 69 (1): 9-15. doi: 10.1080/00029890.1962.11989827
- [7] Hylland A, Zeckhauser RJ. The efficient allocation of individuals to positions. *Journal of Political Economy* 1979; 87 (2): 293-314. doi: 10.1086/260757
- [8] Teo CP, Sethuraman J. The geometry of fractional stable matchings and its applications. *Mathematics of Operations Research* 1998; 23 (4): 874-891. doi: 10.1287/moor.23.4.874
- [9] Roth AE, Sonmez T, Unver MU. Kidney exchange. *Quarterly Journal of Economics* 2004; 119 (2): 457-488. doi: 10.1162/0033553041382157
- [10] Roth AE. The evolution of the labor market for medical interns and residents: a case study in game theory. *Journal of Political Economy* 1984; 92 (6): 991-1016. doi: 10.1086/261272
- [11] Roth AE. The economist as engineer: game theory, experimentation, and computation as tools for design economics. *Econometrica* 2002; 70 (4): 1341-1378. doi: 10.1111/1468-0262.00335
- [12] Manlove DF. Hospitals/Residents problem. In: Kao MY (editor). *Encyclopedia of Algorithms*. New York, NY, USA: Springer US, 2008, pp. 926-930.
- [13] Adany R, Kraus S, Ordonez F. Allocation algorithms for personal TV advertisements. *Multimedia Systems* 2013; 19 (2): 79-93. doi: 10.1007/s00530-012-0284-y
- [14] Chen Y, Berkhin P, Anderson B, Devanur NR. Real-time bidding algorithms for performance-based display ad allocation. In: 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; San Diego, CA, USA; 2011. pp. 1307-1315.
- [15] Zhang P, Bao Z, Li Y, Li G, Zhang Y et al. Trajectory-driven influential billboard placement. In: 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining; London, UK; 2018. pp. 2748-2757.
- [16] Huang M, Fang Z, Xiong S, Zhang T. Interest-driven outdoor advertising display location selection using mobile phone data. *IEEE Access* 2019; 7: 30878-30889. doi: 10.1109/ACCESS.2019.2903277
- [17] Naik PA, Mantrala MK, Sawyer AG. Planning media schedules in the presence of dynamic advertising quality. *Marketing Science* 1998; 17 (3): 214-235. doi: 10.1287/mksc.17.3.214
- [18] Cosgun O, Gultas I, Serarslan MN. Application of a mathematical model to an advertisement reservation problem. *An International Journal of Optimization and Control: Theories & Applications (IJOCTA)* 2012; 3 (1): 23-33. doi: 10.11121/ijocta.01.2013.00133
- [19] Zhang X. Mathematical models for the television advertising allocation problem. *International Journal of Operational Research* 2006; 1 (3): 302-322. doi: 10.1504/IJOR.2006.009303
- [20] Turner J. *Ad slotting and pricing: new media planning models for new media*. PhD, Carnegie Mellon University, Pittsburgh, PA, USA, 2010.
- [21] Turner J, Scheller-Wolf A, Tayur S. OR PRACTICE—scheduling of dynamic in-game advertising. *Operations Research* 2011; 59 (1): 1-16. doi: 10.1287/opre.1100.0852
- [22] Zhao P. *Collaborative real-time interactive billboard networks*. MSc, University of Waterloo, Waterloo, Ontario, Canada, 2015.
- [23] Nautiyal A, McCabe K, Hossari M, Dev S, Nicholson M et al. An advert creation system for next-gen publicity. In: *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*; Dublin, Ireland; 2018. pp. 663-667.

- [24] Cetin E, Esen ST. A weapon–target assignment approach to media allocation. *Applied Mathematics and Computation* 2006; 175 (2): 1266-1275. doi: <https://doi.org/10.1016/j.amc.2005.08.041>
- [25] Keskinturk T, Cetin E. A genetic algorithm metaheuristic for the weapon-target based media allocation problem. *Alphanumeric Journal* 2015; 3 (1): 1-6. doi: [10.17093/aj.2015.3.1.5000128271](https://doi.org/10.17093/aj.2015.3.1.5000128271)