

BIN-CT: Urban Waste Collection based on Predicting the Container Fill Level

Javier Ferrer*, Enrique Alba*

Dept. Lenguajes y Ciencias de la Computación, University of Malaga, Spain

Abstract

The fast demographic growth, together with the population concentration in cities and the increasing amount of daily waste, are factors that are pushing to the limit the ability of waste assimilation by Nature. Therefore, we need technological means to optimally manage of the waste collection process, which represents 70% of the operational cost in waste treatment. In this article, we present a free intelligent software system called BIN-CT (BIN for the CiTy), based on computational learning algorithms, which plans the best routes for waste collection supported by past (historical) and future (predictions) data.

The objective of the system is to reduction the cost of the waste collection service minimizing the distance traveled by a truck to collect the waste from a container, thereby reducing the fuel consumption. At the same time the quality of service for the citizen is increased, avoiding the annoying overflows of containers thanks to the accurate fill-level predictions given by BIN-CT. In this article we show the features of our software system, illustrating its operation with a real case study of a Spanish city. We conclude that the use of BIN-CT avoids unnecessary trips to containers, reduces the distance traveled to collect a container by 20%, and generates routes 33.2% shorter than the routes used by the company. Therefore it enables a considerable reduction of total costs and harmful emissions thrown up into the atmosphere.

Keywords: Waste Collection, Machine Learning, Recycling, Forecasting, Routes Generation

1. Introduction

Waste generation is globally growing at an unstoppable rate, in contrast, the assimilation capacity of our planet is decreasing, turning waste treatment into a complex challenge we need to address as a modern society (Al-Salem et al., 2009). In addition, the linear structure of our economy (extract, manufacture, use and disposal) has reached its limits; the Earth is exhausted, natural resources are drained. The unsustainable development of most nations has created a problem due to the growing generation of waste that must be solved. Despite the social and governmental commitment, there are hardly any technological means to optimise the management of the waste collection process. The largest resources are needed to collect the municipal waste, which is mainly created by household, industrial, and commercial activity.

Traditionally solid waste collection was carried out without previously analyzing the demand or the routes of the vehicles and the decisions about the routes were made by the drivers, although the solutions were far from optimal. Solid waste collection is usually undertaken based on static plans, with a pre-determined number of trips per week, designed manually in most cases. This approach has severe limitations due to the number of constraints to consider when finding an optimal solution. Moreover, waste collection planning problems are influenced by stochastic waste generation, traffic conditions, and

many constraints, which are unmanageable by a person, so the use of an automatic tool to solve these types of problems is mandatory.

In this article we propose an intelligent software system, called BIN-CT, for the management of solid waste collection in an urban area. The system is doubly intelligent, since it integrates algorithms to predict the fill level of waste containers, plus the subsequent optimal generation of routes for collection trucks. In other words, our software system solves the two major problems that a waste collection service face: 1) what containers should be collected and 2) in what order should the containers be visited to minimize the cost.

In order to solve the first question, we will use computational machine learning techniques to estimate the fill level of each container. Supported by historical data (past data), the system generates predictions (future data) to decide when a container should be collected. This prediction could be performed long term or short term, considering seasonality, weekends, and holidays. We want to highlight that, as far as we know, the fill-level prediction of all containers individually (fine grain) is a feature, which until now, has not been implemented in any tool for waste collection management. Additionally, our software system is able to interact with an Internet of Things (IoT) system to obtain the actual filling of the containers equipped with volumetric sensors. Moreover, the sensors' information will be essential in the near future to validate the historical data (obtained by truck drivers).

Route generation (of waste collection in our case) is a well-

*Corresponding author

Email addresses: ferrer@lcc.uma.es (Javier Ferrer),
eat@lcc.uma.es (Enrique Alba)

known combinatorial problem called the Vehicle Routing Problem (VRP) (Dantzig and Ramser, 1959). However the use of historical data, predictions and the intelligent decisions about the inclusion/exclusion of containers mean that it is not a simple VRP. The solutions to our problem are usually constrained by many factors. Some of them are the number of available trucks, the trucks' capacity, the amount of waste to collect and the characteristics of the street where the container is placed (too narrow for a big truck), among others. The planning of waste collection routes has a wide number of variants and constraints which makes it unmanageable for one person alone. In contrast, there is great room for improvement using automatic algorithms to deal with this problem. The improvement results in lower cost services, a reduction of harmful emissions and a better service for citizens, in quality and costs.

In addition, we must take into account that the current city system itself is under discussion. Many local authorities have been forced to examine its cost-effectiveness and environmental impact since we have a system based on static collection frequencies (once a week, twice a week, daily...) that sometimes results in unnecessary trips to semi-empty containers. The pollution generated by these trips could be more dangerous for the environment than the benefits of the collection. This is especially critical in the case of selective collection (plastic, paper, glass...), where the waste volume is smaller than in the case of organic waste. So, when you are dealing with recyclable waste, the planning of optimal collection routes is even more influential. Presently, the recyclable waste collection process represents 70% of the operational cost of waste treatment (Teixeira et al., 2004). The main objective of the work presented here is the reduction of the collection process cost and is essentially dependent on the distance travelled by the collection trucks.

In Figure 1 we show the steps followed to solve the problem we face in this article. The main phases are as follows: First, all the data needed to solve the problem is treated and stored in a database. In this way, the BIN-CT system is able to load the information from the database instead of using several large files. In the second phase, the software system uses the historical fill-level data of each container and generates the fill-level predictions. In the third step, the software system decides which containers are going to be included in the next routes. The decision is taken according to a given criterion based on the predictions and estimated fill levels. In the fourth phase the route algorithm computes optimal solutions, i.e., the complete routes that the trucks must follow to collect the containers selected in the previous step. Finally, the best solution is shown on a navigable map in a web browser.

The main contributions of this article are as follows:

- We propose an intelligent system called BIN-CT for generating optimal waste collection routes.
- We compare three machine-learning algorithms for the forecasting of filling levels of each container (fine grain) from historical data.
- We use a new real case study of a city with 217 containers

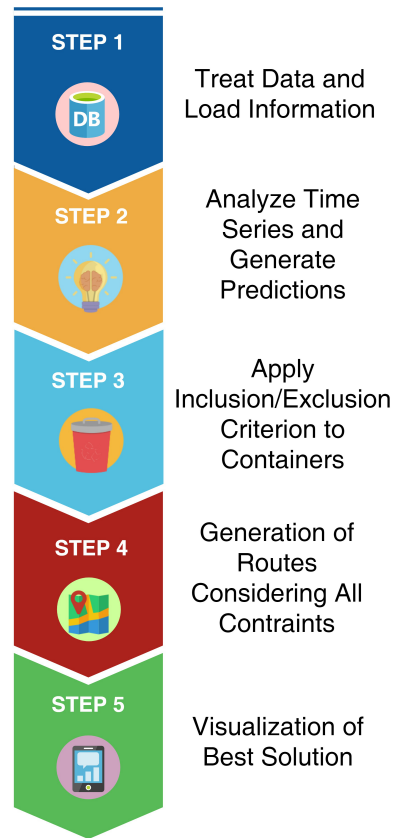


Figure 1: Methodology followed to solve the waste collection problem.

to illustrate the behavior of the proposed system.

- We compare the results obtained by BIN-CT to the results of a real company in a real-world case study.

The remainder of the article is as follows. In Section 2 we describe work related to waste collection optimization, especially the different variants studied in the literature. Section 3 describes the waste collection problem we address in this article. Section 4 explains our proposal, the BIN-CT software system. The four modules that comprise our software system are described and we outline the most important features of the system. Our experimental setup is presented in Section 5, then we describe the real-world case study, and analyze the results of two different experiments. Finally, we present our conclusions and future work in Section 6.

2. Related Work

Waste collection is a process with countless variants and constraints which have led to a multitude of studies in recent years due to its importance. The approaches in the literature could be classified, among other ways, according to the waste type that is treated. Some of them are: *residential waste* commonly known as garbage (Garvin et al., 2011), *industrial waste* where customers/containers are more dispersed and the amount of waste is higher (Rabbani et al., 2018), *recyclable waste* (Dat et al.,

2012) increasingly important for our society, where the collection frequency is lower than organic waste and *hazardous waste* where the probability of harm and costs are both minimized (Rabbani et al., 2018).

In municipal solid waste collection (Beliën et al., 2014), the authorities need global studies to quantify the waste generated in a period of time to be able to manage it. For example, waste generation forecasting for the inhabitants of Xiamen City (China) was studied by (Xu et al., 2013). The main difference with our approach is the granularity of the object under study. They predict the amount of waste produced in the whole city, in contrast, we generate predictions for every single container in a city.

This supposes a considerable increase in the complexity of the problem that is to be solved, because it is necessary to consider multiple aspects such as the location, the customs of the citizens, the population density of the area, etc. Along the same research lines, the impact of the intervention of local authorities on waste collection has also been studied (Cole et al., 2014), with this being relevant in the medium-long term.

Regarding the location where the collection takes place, there exist multiple variants of the problem. There are *communal collections* where the local authority identifies a place shared by the community (Tung and Pinnoi, 2000), in most cases a local waste facility for recycling. On the other side, we found the *kerbside collection* (Sniezek and Bodin, 2006) where the household waste is collected from individual small containers located near each house. The intermediate case studied here is the collection of containers which serve several streets and blocks of flats (Sakti et al., 2018).

In medium/large cities, the number of containers is in the range of hundreds or even thousands, therefore the use of advanced computational methods are required to find an optimal solution. In the past, exact methods were used to solve this problem like *Branch and Bound*, based on *mathematical programming* (Arribas et al., 2010). However the execution time tends to increase exponentially with the number of containers, so the use of hybrid methods like matheuristics Gruler et al. (2017a) and bioinspired methods is recommended. Although some metaheuristics have already been explored in the literature such as *Ant Colony optimization*, *Genetic Algorithms* (Buenrostro-Delgado et al., 2015), *Greedy Randomized Adaptive Search Procedure* Gruler et al. (2017b), or *Variable Neighborhood Search* Gruler et al. (2017c). There is still room for improvement when one is dealing with the waste collection problem at the fine granularity we address it in our approach, that is, analyzing each of every single container.

The waste collection process implies a large quantity of stakeholders like authorities, citizens, the company in charge of the collection service and the company workers. Depending on the stakeholder, the objectives to optimize are different. Some examples of optimization goals are the *number of vehicles* for the service (Ombuki-Berman et al., 2007), the service *total cost* (Arribas et al., 2010), the *environmental impact* (Tavares et al., 2009), the *required staff* for the service (Hansmann and Zimmermann, 2009), the collection *route length* (Ustundag and Cevikcan, 2008), and the *total time* of the waste collection (Ar-

ribas et al., 2010). Our approach considers several objectives at a time, finding solutions which minimize the number of vehicles needed, the total cost, the route length, and the environmental impact.

3. Problem Description

Waste collection management is a global problem that affects most cities around the world. In fact, this problem falls within the initiative of Horizon 2020 and Smart City for innovation and research, promoted by the European Union. The reduction of energy consumption and the responsible use of resources has become key to combat the rigors of the economic crisis. Specifically, one of the objectives of the Smart City initiative is to reduce the emission of greenhouse gases, use sustainable resources, and efficiently manage energy sources. We are therefore dealing with a problem of general interest in this article. We focus on some of the aspects discussed in Section 2. Specifically, we study recyclable waste, we use fine grain predictions for every single container, we manage hundreds of containers, and we use an evolutionary algorithm as resolution technique.

The problem consists of planning how the waste collection will take place in a determined area. The objective is to define optimal collection routes for the available trucks whose capacity cannot be exceeded. An optimal collection route is an ordered list of collection places to visit that minimizes the total distance traveled by all the trucks involved.

Since the number of containers is great for all of them to be collected every day and it is not a cost-effective strategy, the containers must be visited at a more appropriate time (ideally when their filling level is close to 100%). In addition, routes are constrained by the trucks' capacities, that may differ. Moreover, there exist constraints which affect specific containers that must be collected by a small truck (with less capacity) due to space restrictions. Route duration is determined by the travel time between containers and the time to empty them.

Generally, there are some difficulties in estimating the distances and duration of trips between containers due to traffic conditions. Most approaches (Nuortio et al., 2006, Teixeira et al., 2004) only use the distance as a criterion for generating the cost matrix. They calculate the shortest paths between pairs of waste containers using the Dijkstra algorithm (Dijkstra, 1959). In contrast, we have opted to calculate the estimated distance and duration between pairs of containers using the Google Maps Api v3. By doing so, we obtain a more accurate value for travel distance and duration taking into account traffic regulations and conditions.

The quantity deposited daily in each container is unknown, it depends on multiple factors such as the number of citizens sharing a container, the seasonality and lifestyle. In other papers, the authors assume the quantity deposited in the containers is constant every day, however, BIN-CT treats each container as a different entity that can change daily. As the previous fill levels of the containers are available, the software system forecasts future fill levels based on historical data. This information will help determine which containers must be collected in the next shift. The predictions generation is really complex, since we

need to model the behavior of the population that deposits its waste in a specific container.

Overall, the complexity of this problem can be derived from the number of restrictions that can be applied to the model, so that the more restrictions applied, the more realistic the solution found and the larger its computational complexity. The constraints we take into account are highly diverse. With respect to the trucks, we can restrict the capacity of each vehicle individually, the number of them and the ability to collect containers in narrow streets. The quality of service defined by the waste management company (imposed by the local authorities in the contract with the company) could be seen as a soft constraint. In this paper the company establishes that all the containers with a filling estimation greater than 80% must be collected. In summary, we enumerate the constraints considered here:

- Containers:
 - Variable number of containers
 - Different demands
 - Different locations
 - Custom unloading time
 - Only specific trucks can unload a particular container
- Vehicles:
 - Variable number of vehicles
 - Different loading capacities
 - Ability to unload specific containers (narrow streets)
 - Custom cost based on distance traveled

The vehicle routing problem is well known to be NP-hard and the waste collection problem described here is basically a VRP with several constraints, so it is even harder to solve. Furthermore, a realistic instance involving 50 or more containers and several vehicles, means thousands of decision variables and constraints Varone et al. (2014). Therefore the use of heuristics to obtain quasi-optimal solutions in a moderate time is recommended. In this paper, we use evolutionary algorithms for this purpose.

3.1. Problem Formulation

The underlying VRP in the waste collection problem is a combinatorial problem whose ground set is the edges of a graph $G(C, A)$. The notation used for this problem formulation is as follows:

- $C = \{c_0, c_1, \dots, c_n, c_{n+1}\}$ is a container set, where a depot (0 denotes the depot) is located at c_0 and c_{n+1} , and $C' = C \setminus \{c_0, c_{n+1}\}$ is the set of n containers.
- $A = \{(c_i, c_j) | c_i, c_j \in C; i \neq j\}$ is a directed arc set.
- D is a matrix of non-negative distances (meters) d_{ij} between containers c_i and c_j .
- T is matrix of non-negative duration (seconds) t_{ij} between containers c_i and c_j .

- R_i is the route for vehicle i .
- V_i is the vehicle i which has a maximum capacity v_i .
- v is the number of vehicles. One route is assigned to each vehicle.
- s is a binary vector, which indicates if V_i has the size to access the location of certain containers (e.g. those located in the city center).
- w is a binary vector, which indicates if the container c_i must be collected by a vehicle $V_i | s_i = 1$.

In a real city the distance between a container i and j could be different due to traffic regulations, therefore when $\exists i, j D_{ij} \neq D_{ji}$ the problem is said to be asymmetric.

Each vertex c_i in C' has an associated quantity q_i of waste to be loaded by a vehicle. The VRP thus consists in determining a set of v vehicle routes of minimal total cost, starting and ending at a depot, such that the maximum waste is collected and a container is visited at most once. Note that each vehicle has an associated maximum load capacity v_i which can not be exceeded, i.e., the total quantity of waste of all containers serviced on a route R_i must not exceed the vehicle capacity v_i .

A feasible solution is composed of a set of routes $R_1, \dots, R_i, \dots, R_v$ where R_i is the route for the vehicle i and a permutation σ_i of $R_i \cup 0$ (0 denotes the depot) specifying the order of the containers on route i . In addition, each route has to fulfill the following constraint related to the size of the containers: if $\exists c_j \in R_i$ such that $w_j = 1$ (e.g. the container is in the city center), then $s_i = 1$ (the vehicle has the size to reach those containers).

Additionally, we consider a service time δ_i (time needed to load the container), required by a vehicle to load the quantity q_i at container c_i . In this context, the time duration γ_i to collect c_j is the service time δ_i plus the travel time t_{ji} . Note that in this approach γ_i (costs associated with the travel time and service time) are used to report an estimation of the duration of the routes.

The cost of a given route ($R_i = \{c_0, c_1, \dots, c_{v+1}\}$), where $c_i \in C$ and $c_0 = c_{v+1} = 0$ (0 denotes the depot), is given by

$$C(R_i) = \sum_{i=0}^v D(c_i, c_{i+1}) \quad (1)$$

Finally, the fitness for the problem solution S is given by:

$$Fitness_S = \sum_{i=1}^v C(R_i) + p * nac \quad (2)$$

where p is a penalty value and nac is the number of unassigned containers (due to vehicles' maximum capacity constraint). In this paper we set $p = 500$, a high enough value for this instance.

4. Our Proposal: BIN-CT Predictive System

The BIN-CT predictive system has been designed to improve municipal waste collection planning. Our software system is

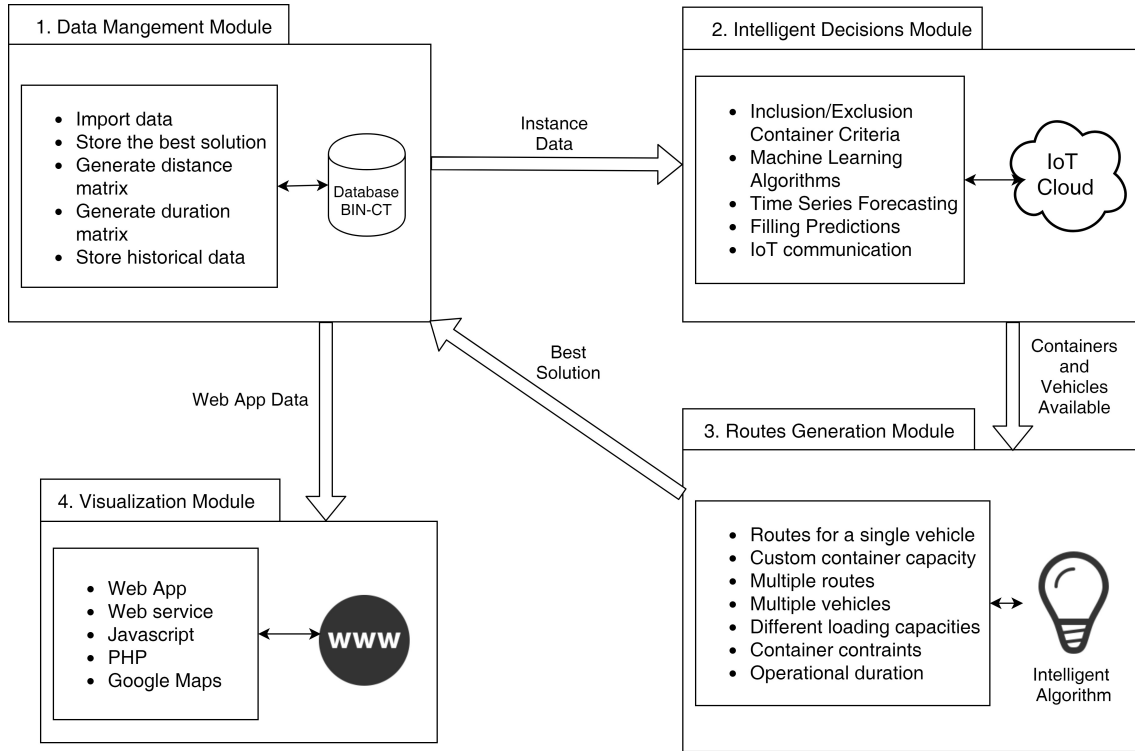


Figure 2: BIN-CT module diagram.

a combination of two main algorithms, one for predicting the filling levels of the containers and one for computing the best routes to visit them. These algorithms are encoded into a software package that offers a Graphical User Interface (GUI) so that a manager in the recycling company can make informed decisions and take action, like giving daily routes to truck drivers to follow or pursuing new offline studies like analyzing the evolution in time of the filling level of the containers.

To deal with the complexity of the global problem, we address it in four sequential phases. In the first phase, the travel distances and duration for all pairs of containers are computed to give more realism to the solution. Then, in the second phase, the fill levels are estimated and the containers which are not going to be part of the routes for the next service are discarded. In the third phase the routes are generated, minimizing the total cost, and in the last phase we show the best solution in a web browser. Therefore, we divided the tool into four different modules which can be seen in Figure 2.

In this section we detail the main features of our software, as well as the solution quality it achieves when solving an instance of the solid waste collection problem described here. The BIN-CT predictive system comprises four software modules: A Data Management module, Intelligent Decision module, Routes Generation module and Visualization module.

4.1. Data Management Module

All information for the execution of the software that comes from the waste management company, which undertakes the collection, can be loaded from a database. The use of a database is desirable due to the amount of information needed to solve

Table 1: Input data for the BIN-CT system.

Entity	Parameters
Containers	Identifier
	Location (latitude, longitude)
	Max Capacity
	Volumetric Sensor (Yes or No)
	Unload Time
	Address
	Group
Historical Data	Small (Yes or No)
	Sensor Identifier
	Container Identifier
	Date
Vehicles	Collected Yesterday (Yes or No)
	Day
	Waste Collected
	Max Capacity
	Small (Yes or No)
	Registration Number
	Costs

the problem. The main entities (input and output data) used in our system are shown in Figure 3. Additionally, we list the input data needed to run BIN-CT effectively in Table 1.

The main entity is *Containers*. We store the identification of the container, information about the location such as coordinates and addresses, the maximum capacity of the container, how long it takes to collect the container, the type of waste

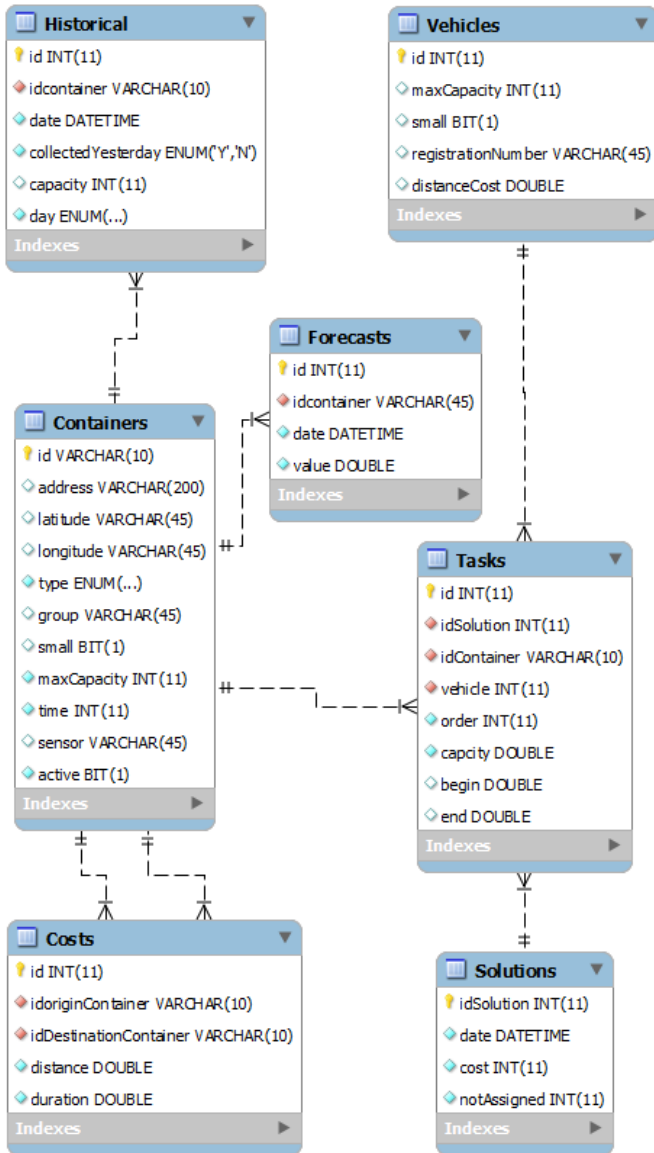


Figure 3: BIN-CT database diagram. We store the information in a clear and structured way in seven tables.

(glass, paper,...), the type of the truck needed to collect the container (e.g., small truck in narrow central streets) and whether it has a sensor or not to measure the fill-level.

The container is related to the historical filling data entity where we store the filling data of each container. It is also related to the *Costs* entity where we store the distance and estimated duration of the trip between containers. This information is extracted with Google Maps Api v3 and is used to solve the underlying VRP. The generation of the duration and distance matrix is essential to generate a realistic solution.

The table *Forecasts* stores the predictions generated by the machine-learning algorithms that are used to estimate which containers will be collected in the next service. To undertake a service, the company has vehicles that have different sizes and loading capacities, which we must consider as constraints. A

smaller truck has less loading capacity; thus it usually collects fewer containers.

Finally, once the execution of the route optimization algorithm has finished, we store the best solution in the database for visualization purposes. The solution of the whole problem is an ordered list of containers that must be visited by each vehicle. This information is stored in the entities *Solution* and *Tasks*.

4.2. Intelligent Decisions Module

The fundamental characteristic of our software is the ability to predict the daily waste contribution for each of the containers that are part of the analyzed set. We have carried out an individualized mathematical modeling of each container using Weka's machine learning library (Frank et al., 2016). Specifically we have used the time series analysis, which is the process of using statistical techniques to model and explain a time-dependent series of data points. Therefore, time-series forecasting is the process of using a model to generate predictions (forecasts) for future events based on known past events, like the fill-level of each container. The time-series forecasting algorithms enable the tool to generate filling predictions for those containers not monitored with sensors.

In this module we have also defined the inclusion/exclusion criteria used to make decisions about the containers that will be collected using the generated routes. The inclusion/exclusion criteria can be defined according to container capacity, forecasts made, and real readings of the sensors, if possible. In addition, given a set of containers and a criterion, we have the possibility of forcing the inclusion of some containers with higher priority on the routes, and some may be excluded due to the defined criteria and constraints.

4.3. Routes Generation Module

This module has, as its main functionality, the generation of optimal routes for the waste collection. Given a set of containers, we use an evolutionary algorithm EA(1+1) based on the principle of ruin and recreation (Schrimpf et al., 2000) that allows us to generate efficient collection routes, taking into account the distance and the driving time between them.

In general, all the containers have the same capacity, but as time goes on, city managers buy new containers with potentially new, different capacities. Our software considers containers of different sizes and, therefore, different capacities. The capacity of each one is managed individually, and this allows us to estimate the total waste that the vehicle will collect, since we cannot exceed the load limit. In addition, vehicle fleets are not only made up of identical vehicles, but vehicle fleets are heterogeneous, so vehicles are also considered individually. For example, we can assign a larger number of containers to vehicles with higher capacity.

Sometimes there are special cases where the compatibility between container and vehicle has to be taken into account. This is the case when a container has to be picked up by a small vehicle. Our software allows us to take this restriction into account in order to generate valid routes. In addition, this feature can also influence the container collection time, which is the

time it takes to the driver to pick up the container once it is in front of it. This input time is assigned depending on the location or characteristics of the container to be collected. This characteristic has a beneficial effect on the fact that the total estimated collection time is more realistic.

4.3.1. Algorithm

The evolutionary algorithm EA(1+1) used to solve the routing problem is based on the ruin and recreate principle. It is a large neighborhood search that combines elements of simulated annealing and threshold-accepting algorithms. This approach is suitable for complex problems that have many constraints (Schrimpf et al., 2000, Misevicius, 2003). It is an all-purpose metaheuristic that can be used to solve a number of classical VRP types and basic search strategies can be easily varied to small and large movements according to the complexity of the problem. The pseudocode of the algorithm is shown in Algorithm 1.

Algorithm 1 Pseudocode of the algorithm.

Inputs: Set of containers (location, fill-level, predictions), set of vehicles, distance matrix, duration matrix, deposit location

Output: Set of ordered lists of containers

```

1: bestSolution ← randomSolution()
2: while evals < totalEvals do
3:   initialSolution ← randomSolution()
4:   removedTasks ← ruin(initialSolution)
5:   newSolution ← reCreation(removedTask)
6:   evaluate(newSolution)
7:   if (newSolution < bestSolution) then
8:     bestSolution ← newSolution
9:   end if
10: end while
11: return bestSolution

```

The algorithm starts with a random initial solution. It divides parts of the solution leading to a set of tasks (collecting a container) that are not going to be assigned to a vehicle and a partial solution containing all other tasks. This process is called ruin the solution because we extract some tasks from the solution. Based on the partial solution, the algorithm reintroduces the tasks extracted leading to a new solution. Hence, this is the recreation process. If the new solution has more quality than the best-so-far solution, it is accepted as the new best solution. These steps are repeated until the algorithm reaches a number of iterations.

4.4. Visualization Module

The solutions are shown in an HTML navigable map. The software generates solutions for one or more days, and these can be viewed in a web browser. In Figure 4 we show the interface of the web application, which can also be viewed online at <http://mallba3.lcc.uma.es/binct/>. The main part shows a solution obtained for two vehicles: one red route and another black route, with two different trucks. In addition, we can observe the containers marked in red, which will be those whose prediction indicates a percentage of filling higher than 80%.

The solutions can be accessed from any mobile device and the user interface adapts itself to the platform on the fly. This feature is especially desirable when the trucks' drivers have any doubts regarding the route they should follow.

5. Experiments

In this paper we carry out two different experiments. First, we focus on the predictions, which is the most innovative part of our system. They will allow us to achieve the greatest savings in the operation of our waste collection service. This first experiment consists in the comparison of three different algorithms to generate predictions.

The second experiment is the generation of routes for a working day, starting from the real state that the containers had, so that, they have different filling states. These fill levels are complemented by the predictions for the next day. We compare our solution with the real solution provided by the company's human experts for that exact day (what really happened that day). In this way, we try to show that our approach is realistic, and therefore our solutions are too. As quality measures we analyze the distance, duration, and amount of waste collected using the generated routes. As a performance measure we report the time required to run the algorithm. We completed 30 independent runs in order to provide a fair value for the runtime. Finally, the experiments were run on a machine with Intel Core i7-3770 processor at 3.40 GHz and 16 GB memory.

5.1. Case of Study

In this article we illustrate the behavior of BIN-CT with a real case study of an Andalusian city (Spain), where we highlight the benefits of our approach, showing it to also be effective and realistic. Our case study considers 217 paper containers from the metropolitan area of the city. The choice of an instance of recycling waste is more attractive than an organic waste collection. In this sense, we show the quality of our approach because most paper containers do not need to be emptied every day like organic waste ones, so they have a high variability in collection frequency. The problem instance parameters used in the case study are shown in Table 2.

Table 2: Case study main parameters.

Parameter	Value
Regular containers	208 units
Small truck containers	9 units
Total containers	217 units
Container capacity	75 kg
Trucks	2 units
Unload container	3.5 min
Small truck capacity	1700 kg
Big truck capacity	2000 kg
Historical data	11 months

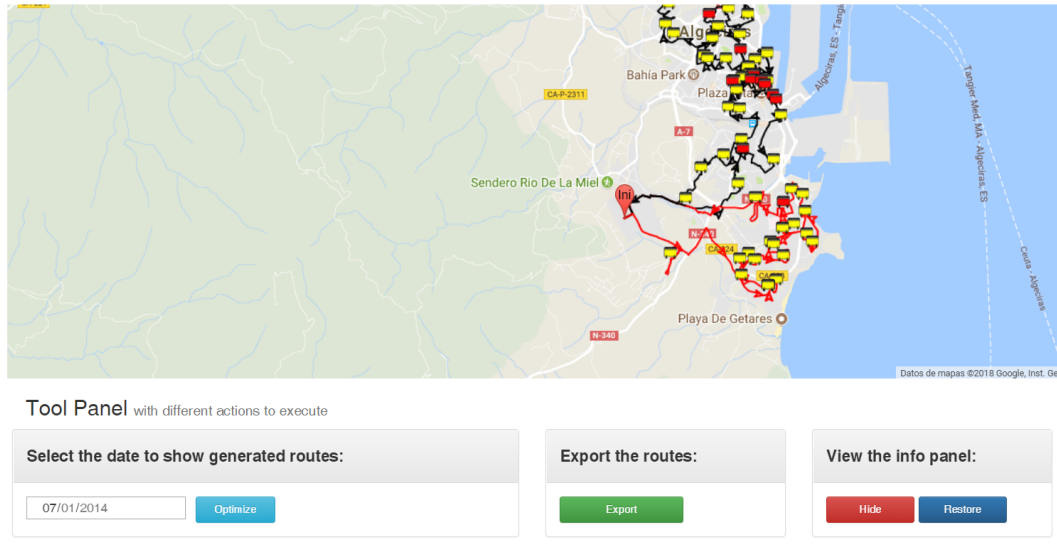


Figure 4: BIN-CT App interface.

5.2. Experiment 1: Comparison of Time Series Algorithms

This experiment is intended to carry out a comparison between three time-series algorithms used to forecast the fill-level for all containers. The algorithms used in this comparison are: Linear Regression (Su et al., 2012), Gaussian processes (Mackay, 1998) and Support Vector Machine for regression (SMOreg) (Rivas-Perea, 2013). We briefly describe the techniques on which the algorithms are based:

- Linear regression works by estimating coefficients for a line that best fits the training data. It is a good idea to evaluate linear regression on a problem as a baseline, before moving onto more complex algorithms.
- Gaussian process is a collection of random variables, any finite number of which have a joint Gaussian distribution. A Gaussian process is completely specified by a mean function and a positive definite covariance function.
- SMOreg works by finding a line of best fit that minimizes the error of a cost function. This is done using an optimization process that only considers those data instances in the training dataset that are closest to the line with the minimum cost.

In this experiment we use the fill-level of all 217 containers in 11 months. The data is obtained by the truck driver when the container is unloaded. If the container is not visited, then there is no datum for that specific day. Thus, the system must deduce what happened on those days for which we do not have data.

In order to alleviate this issue, we have performed a pre-processing of the data that has allowed us to obtain the estimated daily load increase of the containers. From the processed data, we have generated the daily predictions for the following month using the three forecasting algorithms.

The results indicate that the learning algorithm based on Gaussian processes is the best algorithm in our comparison, obtaining an average of just 3.83% mean absolute error in the filling predictions of next month. Regarding the other techniques, the algorithm based on linear regression has an average of 7.41% mean absolute error and the SMOreg algorithm has 9.52%. After performing the Kruskal-Wallis statistical test (with Bonferroni correction to compare more than two samples) with 95% confidence, we can state that there are significant differences between the results of the algorithm based on Gaussian processes with the other two algorithms of the comparison.

As the main conclusion of the first experiment, we can infer that the estimation provided by the best algorithm is more than acceptable, with less than 4% mean absolute error. Note that a container is scheduled for the next collection service when the filling estimation is greater than 80%, so the error is so negligible as to not affect the decision to collect a container. After analyzing these results, we decided to use the algorithm based on Gaussian processes for the second experiment.

5.3. Experiment 2: The Daily Use of the System

In the second experiment, we want to illustrate the general behavior of the system considering all the parameters specified in Table 2. In this experiment we generate routes for the first working day of the twelfth month starting from the real state of the containers and then we compare our solution with the real solution implemented that day. Once our system has given the predictions for the following day, we obtain the estimated fill-level for all containers. Next, BIN-CT chooses the subset of containers that fulfills the following inclusion/exclusion criteria. Those containers with a fill-level higher than 50 % are considered optional, while those with a filling greater than 80 % are marked for mandatory collection. The application of this criterion to the set of 217 containers results in a subset of 77

Table 3: Comparison between the real solution and the solution generated by our system. We highlight the best result for each indicator.

Solution	Truck	Containers	Duration (s)	Distance (m)	Collected (kg)
Experts	Small	31	12,044.00	34,554.00	1,665.00
	Big	31	13,912.00	37,799.00	1,838.00
	Total	62	25,956.00	72,353.00	3,503.00
	Average	-	418.64	1,166.98	56.50
BIN-CT	Small	33	13,191.00	32,407.00	1,689.00
	Big	41	13,960.00	25,281.00	1,984.00
	Total	74	27,151.00	57,688.00	3,673.00
	Average	-	366.90	779.57	49.63

containers to be picked up by two vehicles (the smallest is able to pick up the nine containers that cannot be collected by the big truck).

In Table 3 we show the containers collected, the duration and distance of the routes, and the load of the trucks in the experts solution and the solution provided by BIN-CT. In addition, we show the totals of the measures for both vehicles and the average per container. After the execution of the route generation algorithm, we obtain two different routes: the biggest truck should pick up 41 containers and the smallest truck should visit 33 containers. Therefore, our solution is able to schedule most of the containers with only 3 containers not scheduled for the next service due to truck capacity constraints.

The system provides very accurate information on the duration of the routes, the duration of the four routes can be seen in the fourth column in Table 3. On the one hand, the route of the small truck has an estimated duration of 3 hours, 39 minutes and 51 seconds to collect a total of 33 containers. On the other hand, the route of the big truck will last 3 hours, 52 minutes and 40 seconds to collect a total of 41 containers. However, if we take a look at the distance, the small truck’s route is 32.4 km compared with just 25.2 km of the other route. Therefore, it seems that the constraint of picking up those 9 containers, which cannot be collected by the big truck, is a detrimental to the small truck’s route. Regarding the trucks load, the amount of waste collected is proportional to each truck’s size.

Let us compare the experts solution with our solution. First of all, we analyze how many containers, which are also in the expert’s solution, were selected by BIN-CT. The result is that 66% of the experts’ solution’s containers are also collected in the solution provided by BIN-CT. This means that using static plans, sometimes they collect containers that are below 50% according to our estimations. Actually, around 20% of the containers (12 units) of the studied day had less than 50% fill-level. Our system is therefore unique in detecting them and saving costs in the planning.

The routes provided by BIN-CT are able to pick up more containers, in less time per container, traveling fewer kilometers and taking more advantage of the load capacity of the trucks. We must highlight the difference in distance traveled. We would save 20% of the distance traveled in comparison with the real solution even though our routes took 19 minutes and 55 seconds more than the real ones. If we used the average distance to collect a container and multiply that by the 62 containers

of the real solution, then the estimated distance traveled would be 48,333 meters, which means we would have 33.2% shorter routes than the routes used by the company. The algorithm runs with the default parameters and a stop criterion of 10,000 iterations. The median runtime was 27,350 milliseconds that is less than half a minute, so the algorithm is quite fast. Moreover, if we increase the iteration budget, potentially we would obtain even better results.

After specialized staff from the company, who gave us the data for the case of study, validated the solutions generated by BIN-CT, we can conclude that our system generates realistic solutions and reproduces with enough reliability, what happens in reality. In addition, the use of the fill predictions discarded 140 containers not included as nodes to visit in the generation of routes process. Actually, all containers (except one) which were not included in our routes but included in the real ones, were discarded based on the forecasting and not because of truck capacity constraints. This fact means that with high probability we have saved unnecessary visits to semi-empty containers that are part of a predefined static route for the day analyzed. Therefore using BIN-CT implies a reduction in both the cost of the routes and the harmful emissions thrown by the trucks.

6. Conclusions and Future Work

Traditional solid waste management is usually based on previous experience rather than data. Currently the waste collection plans are static with a predetermined number of visits per week and designed manually, which implies a slow adaptation to any change in normal conditions. In this paper we have proposed BIN-CT, an intelligent predictive software system designed to use knowledge extracted from historical data and technology that is part of the state-of-the-art in urban sustainability research. We are convinced that our intelligent system will help to improve the quality of life of the city where it is implemented. The reduction of costs and the efficient treatment of the waste generated will constitute a way to achieve sustainability in the city.

Our aim is to reduce the costs of the waste collection service, while increasing the quality of service for the citizen. Our software suggests the collection of containers above 80 % and thus avoids the annoying overflows of containers. Through the study case that we have analyzed in this article, we prove that the solutions generated are realistic. At the same time, they are more

efficient because the solutions prevent trips to semi-empty containers and the generated routes are 33.2% shorter than those used by the company. Moreover, after the comparison of three different algorithms, we used the best which generates predictions with just a 3.83% mean absolute error. This fact indicates that the forecasting performed is fairly reliable for the calculations we need to make in our system.

BIN-CT has a direct application for waste management companies, both in Spanish cities and elsewhere in the world. This non-commercial software system, based on computational learning algorithms, does not require a large investment in infrastructure, which is why it is highly attractive to companies. There is also an interesting scientific side (use of time series, predictions, efficient algorithms, ...) that companies may be interested in, which could make the leap in quality and distinction that is sought after in a market as competitive as the current one.

As for future work, we highlight the possible integration of an indeterminate number of sensors, in a larger cyber-physical system, to measure the amount of waste in the containers. The way the system is built facilitates an easy integration of sensors. The use of sensors makes it possible to change a route in real time by adding or removing containers. Fortunately, this valuable feature could save more unnecessary trips to containers, while at the same time preventing possible overflows of containers.

Regarding the algorithms for the generation of predictions, here we have studied some techniques for time-series prediction based on regression. However, recently the recurring neural network technique has emerged as a good method for the generation of time-series predictions (LeCun et al., 2015, Camero et al., 2018). We plan to improve the fill-level predictions by using an algorithm based on recurring neural networks.

Acknowledgments

This research has been partially funded by the Spanish Ministry of Science and Innovation and FEDER under contracts RTC-2017-6714-5 and TIN2017-88213-R and the network of smart cities CI-RTI (TIN2016-81766-REDT). The CELTIC project C2017/2-2 under contracts #8.06/5.47.4997 y #8.06/5.47.4996. J. Ferrer thanks University of Málaga for his postdoc fellowship.

References

Al-Salem, S. M., Lettieri, P., Baeyens, J., oct 2009. Recycling and recovery routes of plastic solid waste (PSW): A review. *Waste Management* 29 (10), 2625–2643.

Arribas, C. A., Blazquez, C. A., Lamas, A., apr 2010. Urban solid waste collection system using mathematical modelling and tools of geographic information systems. *Waste Management & Research* 28 (4), 355–363.

Beliën, J., De Boeck, L., Van Ackere, J., feb 2014. Municipal Solid Waste Collection and Management Problems: A Literature Review. *Transportation Science* 48 (1), 78–102.

Buenrostro-Delgado, O., Ortega-Rodriguez, J. M., Clemitshaw, K. C., González-Razo, C., Hernández-Paniagua, I. Y., 2015. Use of genetic algorithms to improve the solid waste collection service in an urban area. *Waste management (New York, N.Y.)* 41, 20–7.

Camero, A., Toutouh, J., Stolfi, D. H., Alba, E., 2018. Evolutionary Deep Learning for Car Park Occupancy Prediction in Smart Cities. In: *International Conference on Learning and Intelligent Optimization*. pp. 386–401.

Cole, C., Quddus, M., Wheatley, A., Osmani, M., Kay, K., feb 2014. The impact of Local Authorities' interventions on household waste collection: a case study approach using time series modelling. *Waste management (New York, N.Y.)* 34 (2), 266–72.

Dantzig, G. B., Ramser, J. H., 1959. The Truck Dispatching Problem. *Management Science* 6 (1), 80–91.

Dat, L. Q., Truc Linh, D. T., Chou, S. Y., Yu, V. F., 2012. Optimizing reverse logistic costs for recycling end-of-life electrical and electronic products. *Expert Systems with Applications* 39 (7), 6380–6387.

Dijkstra, E. W., 1959. A note on two problems in connexion with graphs. *Numerische Mathematik* 1 (1), 269–271.

Frank, E., Hall, M. A., Witten, I. H., 2016. The WEKA Workbench.

Garvin, B. J., Cohen, M., Dwyer, M. B., 2011. Evaluating improvements to a meta-heuristic search for constrained interaction testing. *Empirical Software Engineering* 16 (1), 61–102.

Gruler, A., Fikar, C., Juan, A. A., Hirsch, P., Contreras-Bolton, C., feb 2017a. Supporting multi-depot and stochastic waste collection management in clustered urban areas via simulation–optimization. *Journal of Simulation* 11 (1), 11–19.

Gruler, A., Fikar, C., Juan, A. A., Hirsch, P., Contreras-Bolton, C., feb 2017b. Supporting multi-depot and stochastic waste collection management in clustered urban areas via simulation–optimization. *Journal of Simulation* 11 (1), 11–19.

Gruler, A., Fikar, C., Juan, A. A., Hirsch, P., Contreras-Bolton, C., feb 2017c. Supporting multi-depot and stochastic waste collection management in clustered urban areas via simulation–optimization. *Journal of Simulation* 11 (1), 11–19.

Hansmann, R. S., Zimmermann, U. T., 2009. Integrated Vehicle Routing and Crew Scheduling (IVRCS) in Waste Management Part I. *Dagstuhl Seminar Proceedings 09261 Models and Algorithms for Optimization in Logistics*, 1–8.

LeCun, Y. A., Bengio, Y., Hinton, G. E., 2015. Deep learning. *Nature* 521 (7553), 436–444.

Mackay, D. J. C., 1998. Introduction to Gaussian processes. *Neural Networks and Machine Learning* 168 (1996), 133–165.

Misevicius, A., 2003. Genetic algorithm hybridized with ruin and recreate procedure: Application to the quadratic assignment problem. In: *Knowledge-Based Systems*. Vol. 16, pp. 261–268.

Nuortio, T., Kytöjoki, J., Niska, H., Bräysy, O., 2006. Improved route planning and scheduling of waste collection and transport. *Expert Systems with Applications* 30 (2), 223–232.

Ombuki-Berman, B. M., Runka, A., Hanshar, F. T., 2007. Waste collection vehicle routing problem with time windows using multi-objective genetic algorithms. *Proceedings of the Third IASTED International Conference on Computational Intelligence*, 91–97.

Rabbani, M., Heidari, R., Farrokhi-Asl, H., Rahimi, N., jan 2018. Using meta-heuristic algorithms to solve a multi-objective industrial hazardous waste location-routing problem considering incompatible waste types. *Journal of Cleaner Production* 170, 227–241.

Rivas-Perea, P., 2013. Support Vector Machines for Regression: A Succinct Review of Large-Scale and Linear Programming Formulations. *International Journal of Intelligence Science* 03 (01), 5–14.

Sakti, S., Yu, V. F., Sopha, B. M., 2018. Heterogeneous Fleet Location Routing Problem for Waste Management: A Case Study of Yogyakarta, Indonesia. In: *International Conference in Management Science and Decision Making*. pp. 1–11.
URL https://repository.ugm.ac.id/274607/1/2018_ICSDM.pdf

Schrumpf, G., Schneider, J., Stamm-Wilbrandt, H., Dueck, G., 2000. Record Breaking Optimization Results Using the Ruin and Recreate Principle. *Journal of Computational Physics* 159 (2), 139–171.

Sniezek, J., Bodin, L., apr 2006. Using mixed integer programming for solving the capacitated arc routing problem with vehicle/site dependencies with an application to the routing of residential sanitation collection vehicles. *Annals of Operations Research* 144 (1), 33–58.

Su, X., Yan, X., Tsai, C.-L., 2012. Linear regression. *Wiley Interdisciplinary Reviews: Computational Statistics* 4 (3), 275–294.

Tavares, G., Zsigraiova, Z., Semiao, V., Carvalho, M., mar 2009. Optimisation of MSW collection routes for minimum fuel consumption using 3D GIS

- modelling. *Waste Management* 29 (3), 1176–1185.
- Teixeira, J., Antunes, A. P., de Sousa, J. P., nov 2004. Recyclable waste collection planning—a case study. *European Journal of Operational Research* 158 (3), 543–554.
- Tung, D. V., Pinnoi, A., 2000. Vehicle routing-scheduling for waste collection in Hanoi. *European Journal of Operational Research* 125 (3), 449–468.
- Ustundag, A., Cevikcan, E., 2008. Vehicle Route Optimization for Rfid Integrated Waste Collection System. *International Journal of Information Technology & Decision Making* 07 (04), 611–625.
- Varone, S., Markov, I., Bierlaire, M., 2014. Vehicle Routing for a Complex Waste Collection Problem Iliya Markov, Ecole Polytechnique Fédérale de Lausanne Vehicle Routing for a Complex Waste Collection Problem. Tech. rep.
- Xu, L., Gao, P., Cui, S., Liu, C., jun 2013. A hybrid procedure for MSW generation forecasting at multiple time scales in Xiamen City, China. *Waste management (New York, N.Y.)* 33 (6), 1324–31.