CILAMCE 2017
Florianópolis - Santa Catarina - Brazil

AN OBSTACLE-AWARE MODEL TO INDOOR NAVIGATION APPLIED TO SUPERMARKET SHOPPING ROUTING

Wagner C. Silva
Paulo E. M. de Almeida

wagnerbhbr@gmail.com
pema@lsi.cefetmg.br

Laboratório de Sistemas Inteligentes / Centro Federal de Educação Tecnológica de Minas Gerais
Av. Amazonas, 7675, 30510-000, Belo Horizonte, Minas Gerais, Brasil

Abstract.

Some smart applications make sense only when the provider knows exactly where the customer is, so indoor localization systems are important to support smart services when outdoor positioning is not possible. Nowadays, the main way to define position over the Earth is the GPS, that is useful for outdoor location positioning, but has strong limitations for indoor services. Urban population expansion brings challenges and demands public places with adequate structure. Accompanying this development, retail establishments grow gradually in scale and in number of products. Then, the customers need services which could find good routes from actual location to points of interest, for instance, to reduce shopping time. This paper proposes an optimization model that uses genetic algorithms (GA) to solve a graph-based routing model. Hierarchical graphs are used to get the indoor map of a retail establishment, and to represent the commodities that make up the shopping list. Then, GA meta-heuristics are adopted to bring out short routes for the shopping cart. The results show that this model is flexible for different retail store layouts and can find good solutions in reasonable time for the supermarket shopping routing problem.

Keywords: *Computational intelligence, optimization, shopping route planning, indoor navigation*

INTRODUCTION

Some smart applications make sense only when the provider knows where the customer is, so indoor localization systems are applicable to support these services. During centuries, the compass was the foremost instrument to navigation and orientation. Nowadays, the main way to define position over the earth is the Global Positioning System (GPS), that is useful for outdoor location positioning but has strong limitations for indoor localization and navigation.

Urban population expansion brings challenges and demands public places with adequate structure and services. Accompanying this development, the retail establishments grow gradually in scale and in number of products, hence the time for shopping usually increases. Despite this, there are many customers that want to buy their goods quickly and they need services, as best route finders, inside these places, to reduce shopping time. To achieve this goal, this paper proposes an indoor navigation system, emphasizing the delivery of value to customers, venue owners and manufacturers.

The remainder of this paper is organized as follows. Section 2 presents the problem description and Section 3 reviews the related works and literature. Section 4 briefly considers the methodology and the proposed model is described in the Section 5. In Section 6, Genetic Algorithms are described in more details, and in Section 7, simulation and experiment results are presented. Finally, in Section 8 some remarkable results and future research directions are discussed.

PROBLEM DESCRIPTION

In one supermarket, most of the customers want to find the desired goods and proceed to checkout more quickly as possible. In these cases and many others indoor localization systems can be useful. The customers need tools that help their to find efficiently the location where the wish products are.

According to Chen et al. (2015), one the problems which is defined in this scenario is known as Supermarket Shopping Route Planning Problem (SSRPP), and can be described as: a consumer who has a grocery list and goes to his favorite supermarket to buy these goods that are located inside it. The consumer needs, from a entrance (FI), to look up all the products of his list (1, 2...., n), to select and to put them in his shopping cart, and finally to go to a checkout point (FF). Therefore, the routes, or possible solutions to the problem, are represented as a vector of type $S_i = (FI, 4, 7, 2, ..., n, FF)$.

This problem is similar, but not the same, to the well known Traveling Salesman Problem (TSP). In SSRPP, initial and final points are fixed and different from each other, while in TSP they are the same and can be any point among the cities considered.

The products are referenced by the identifiers, and each one these products are located in a slot of some gondola inside the establishment and their location is known on Cartesian plane, according to the model displayed next. The distance between two points is given by $d(p_i, p_{i+1})$ which is used in defining the cost function that measure the total displacement of the shopping cart on each route, according to Eq. (1).

$$minf(x) = \sum_{i=0}^{n-1} d(p_i, p_{i+1}) \tag{1}$$

RELATED WORKS AND LITERATURE REVIEW

The research subjects included applications for indoor localization and navigation, including best route finders, guidance and tracking by wireless technologies. There is also interest for works that deal with the smart supermarket applications and special attention is given for the Genetic Algorithms.

Related Works

In Kholod (2010), the researcher investigated how the relationship between the distance traveled in the route and the total purchase value. The author found a Pearson correlation coefficient ($\rho\ of\ Pearson$) equal to 0.8457, which means that these two quantities have a strong linear dependence. Another interesting fact in this study is that a consolidation was done in the collected database and resulted in an average ticket of 3525 yen (approximately $R\$100,00$ today) and 20 items per order on average. It is interesting to note that other groups are carrying out research on this interesting relationship between the characteristics of a traveled route and the behavior of customers in retail stores (Kaneko, 2016 and Syaekhoni, 2016).

According to Orel (2013), the World market of supermarkets is looking to an increase in the use of self-service technologies during the purchasing process. This study evaluates the level of service, which have a strong relationship with customer satisfaction, mainly self-checkout systems (SCS), and the impact in satisfaction and customer loyalty inside supermarkets and grocery stores in an emerging market (in this case, Turkey).

In a recent investigation, Chen et al. (2015) proposed a GA to solve the planning problem of shopping route in supermarket, which explores the search space seeking to minimize the objective function, the traveled distance. The algorithm uses as input a product list containing their coordinates (x, y) on Cartesian plane, and builds the distance matrix between the points of each products for future use in the fitness function. However, this obstacle-unaware model uses the Euclidean metric to measure the distance, and does not deal with gondolas, freezers, and other supermarket furniture.

Genetic Algorithms

In the evolutionary approach, the algorithms have heuristics based on the theory of evolution and GA represent an important approach in this field. According to Gaspar-Cunha (2013), GA are research methods inspired by the principles of natural selection and genetics.

Encoding

The problems where an essential aim is to find an ideal sequence for a set of points, as the case of SSRPP, a permutation representation seems to be the best alternative. The ordering of the genes in the chromosome is relevant because it represents the purchase order of the products in the list. Another important characteristic of this type of problem is that the chromosome can not contain redundant information (go through an item more than once is not expected). These two characteristics are the main constraints of the permutation class.

One of the advantages of this representation is its ability of not accepting invalid solutions, because of the existence of an algorithm to validate solutions, ensuring that all the items of a purchase list must be in the solution vector, without repetition.

Initial Population and evolutionary strategies

In this study, the population is generated in hybrid way, with 90% of individuals coming from random approaches, with a restriction to create only healthy individuals, who follow the restrictions of the problem, and 10% through a greedy search method. The new population is inserted with the current and the ordering by the fitness is made and the individuals with position grater than the maximum size of the population are discarded.

Fitness Function

The function that measures the level of quality of the solutions is directly related to the problem in question. As mentioned in the definition of the problem, the exploration at search space consists of minimizing the function, according to Eq. (1).

Selection method

The method of selection by tournament will be used, it was proposed later to the work of Holland, and is already consolidated in the scientific community. A number k (tournament size parameter) of individuals are randomly selected and the the element with the best fit is selected. This process is carried out as often as necessary until all the progenitors are selected to give origin the new generation. The k parameter will be defined in a following parameter list, please see Table 4 on Section 5, Experiments.

Genetic Operators

Genetic operators act on progenitors with well-defined and standardized operations to give life to each individual of the new generation. This operations are intimately linked to the representation adopted in the problem that is working in order to guarantee the integrity of the new individuals. This operators work in the progenitors in order to generate new healthy individuals and that contain genetic material of their progenitors. In this study, the focus will be in permutation problem class where the most applicable aspects are the absolute position of the items, precedence and adjacency relations.

Crossover operators are the functions that, given the individuals selected from the current population, perform a cross-pairing by generating new individuals to the next generation. These new individuals carry "genetic material" from their progenitors. The parameter that defines the probability of applying the recombination in a pair of progenitors is Pcr.

The Partially-Mapped Crossover (PMX) operator was used as the recombination operator in this study. According to Goldberg (1985), applied in permutation problems, this operator aims to allow GA to search for better solutions while remaining searching in combinations of alleles. An example of the PMX working is presented next.

In the representation below, there are Parents, P_1 and P_2, with 10 elements each, indexes from 0 to 9 and it has the cut interval, from index $3\ to$ 7, and the child C_1 raised with the genes of P_1 contained in that interval. The values marked with red (3,5,6) in P_2 are already present on the child and the values marked with green (4,7) are not present and only these values, from the cut interval, will be used to swap with C_1. The next process is finding the positions in C_1 to insert theses values.

$$\begin{array}{lccccccccccc}
P_1 & 8 & 4 & 7 & 3 & 6 & 2 & 5 & 1 & 9 & 0 \\
P_2 & 0 & 1 & 2 & \boxed{3} & \boxed{4} & \boxed{5} & \boxed{6} & \boxed{7} & 8 & 9 \\
C_1 & _ & _ & _ & 3 & 6 & 2 & 5 & 1 & _ & _
\end{array}$$

The first is **number 4**, and the corresponding in P_1 is **6**, this value in the P_2 is in the index *6* and it is in the cut interval, so restart the first step using the number 6. The corresponding number of 6 in P_1 is **5**, this value in the P_2 is in the index *5* and it is also in the cut interval, so restart the first step using the number 5. The corresponding number of 5 in P_1 is **2**, this value in the P_2 is in the index *2* and it is not in the cut interval, so insert the **number 4** in C_1 in this position.

$$\begin{array}{lccccccccccc}
P_1 & 8 & 4 & 7 & 3 & 6 & 2 & 5 & 1 & 9 & 0 \\
P_2 & 0 & 1 & 2 & 3 & 4 & 5 & 6 & \boxed{7} & 8 & 9 \\
C_1 & _ & _ & 4 & 3 & 6 & 2 & 5 & 1 & _ & _
\end{array}$$

Now, restart the process to the next, **number 7**, and it will produce this result above:

$$\begin{array}{lccccccccccc}
P_1 & 8 & 4 & 7 & 3 & 6 & 2 & 5 & 1 & 9 & 0 \\
P_2 & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\
C_1 & _ & 7 & 4 & 3 & 6 & 2 & 5 & 1 & _ & _
\end{array}$$

Finally, copy any remaining positions from P_2 to the child and it must result in:

$$\begin{array}{lccccccccccc}
P_1 & 8 & 4 & 7 & 3 & 6 & 2 & 5 & 1 & 9 & 0 \\
P_2 & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\
C_1 & 0 & 7 & 4 & 3 & 6 & 2 & 5 & 1 & 8 & 9
\end{array}$$

To bring up the next child C_2, is necessary only to repeat the same process, swapping parent 1 and parent 2 in the begin.

On the other hand, the mutation operators act on the individuals originated by the recombination functions and perform transformations in own genetic material to generate small perturbations in order to avoid stagnation of the solutions in a single region. Such agitation can not guarantee an improvement in the solutions, however it is beneficial to exploring new regions of the search space. The parameter which determine the probability of applying the mutation in a given individual is Pmt.

The mutation operators used in this study consists to select two random positions in the solution vector and perform the exchange of values. For example, if the chromosome is $S_i = \begin{bmatrix} 5 & \boxed{2} & 7 & 3 & 1 & 4 & \boxed{6} \end{bmatrix}$, whit marked positions randomly selected, the new individual must be $S_i = \begin{bmatrix} 5 & \boxed{6} & 7 & 3 & 1 & 4 & \boxed{2} \end{bmatrix}$ after performing a swap operation.

MODEL

This Section introduces a mathematical and optimization model to solve SSRPP problem. All measurements were performed and are presented in meters. The grocery store has several shelves to display products and these also represent obstacles to the movement of people and shopping carts. Therefore, the distance between the points of interest in this problem can not be calculated using Euclidean distance and the Manhattan Distance concept, where the distance between the two points is given by the sum of the absolute differences of their coordinates, is used.

The key premises of this study are: (a) Each gondola is divided into slots of symmetrical size, but may vary between different gondolas; (b) The localization of product is known (gondola and slot). Some model simplifications are made, to achieve the current model: (1) Establishment has symmetrical measures (rectangular or square); (2) The desired precision is 0.5 meters.

Establishment layout and graph model are presented by Fig. 1. The floor plan is shown with the layout of gondolas in retail store, on the left side, and on the right, the model representation, a static graph example, to see relationship between shelves and the graph nodes. In the next Section, a shelf layout, generated by the algorithm, similar to this model will be shown. It will have details of the measures of establishment, including gondolas, slots and products position.

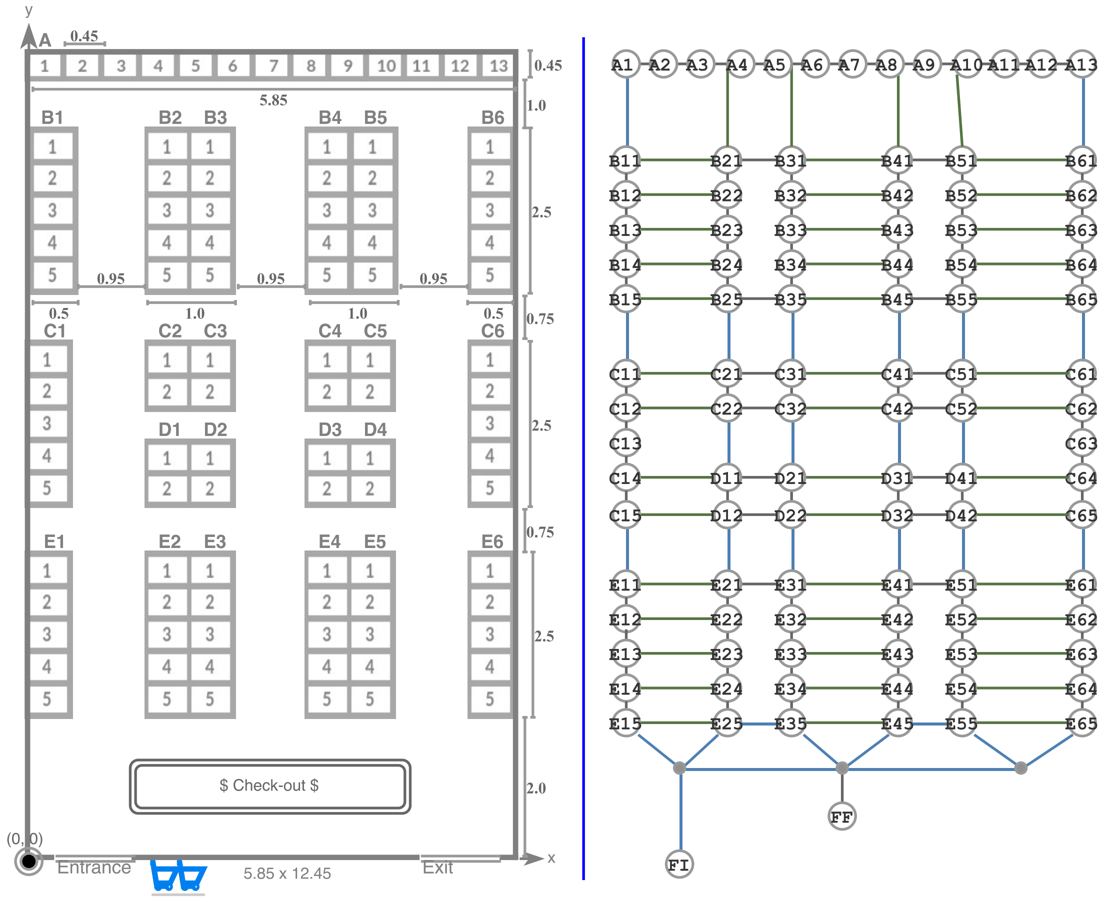

Figure 1: Supermarket layout

EXPERIMENTS

The Supermarket layout is defined by some key information such as supermarket dimensions, location and measurement of gondolas, size and number of slots. The basic information of the establishment layout is shown in Table 1.

The details about the shelves of the retail are presented in Table 2. These data are used to generate a model that will be the input to get the distance matrix. This data structure was chosen because it is simple to be extracted from the back office database system of the retail establishment, by means of a text file, for instance.

Table 1: Retail store basic data

Branch ID	Width	Length	Initial Point	Final Point
1	5.6	12.5	(1.0, 0)	(2.92, 1.0)

Table 2: Shelves layout table

Shelf Id	Width	Length	XCoord	YCoord	Slot Len
A	5.85	0.45	0.00	12.5	0.45
B1	0.50	2.50	0.00	11.0	0.50
B2	0.50	2.50	1.45	11.0	0.50
B3	0.50	2.50	1.95	11.0	0.50
B4	0.50	2.50	3.40	11.0	0.50
B5	0.50	2.50	3.90	11.0	0.50
B6	0.50	2.50	5.35	11.0	0.50
C1	0.50	2.50	0.00	7.75	0.50
C2	0.50	1.00	1.45	7.75	0.50
C3	0.50	1.00	1.95	7.75	0.50
C4	0.50	1.00	3.40	7.75	0.50
C5	0.50	1.00	3.90	7.75	0.50
C6	0.50	2.50	5.35	7.75	0.50
D1	0.50	1.00	1.45	6.25	0.50
D2	0.50	1.00	1.95	6.25	0.50
D3	0.50	1.00	3.40	6.25	0.50
D4	0.50	1.00	3.90	6.25	0.50
E1	0.50	2.50	0.00	4.50	0.50
E2	0.50	2.50	1.45	4.50	0.50
E3	0.50	2.50	1.95	4.50	0.50
E4	0.50	2.50	3.40	4.50	0.50
E5	0.50	2.50	3.90	4.50	0.50
E6	0.50	2.50	5.35	4.50	0.50

As discussed, the product location is known so that, for each product, there are at least information about: Shelf Id, Slot Id and Product Id, as the example in Table 3.

The distance matrix is calculated from information of these tables according to the products that are into shopping list and their positions in the grocery gondolas. This matrix is the key input for the optimization model.

The algorithm flow consists of: (1) To load text files and make instances; (2) To build a graph model with definition of location for each product; (3) To build a distance matrix; (4) To get a shopping list; (5) To get the instance matrix for the shopping list; and (6) To run GA and

Table 3: Commodity location data

Shelf Id	Slot Id	Prod Id	Prod Desc
A	1	1	White Rice
A	1	2	Japanese Rice
A	2	3	Beans
A	9	4	Pasta
B1	3	5	Coffee
B4	2	6	Juice
C2	1	7	Patato
C6	3	8	Cheese
D2	1	9	Tomato
D3	2	10	Salmon
E1	1	11	Beer
E5	5	12	Bread
B1	3	13	Suggar
E2	3	14	Tea

collect the best found shopping route.

For the complete definition of GA, besides the items previously discussed, there is an important set of parameters that directly influence their behavior. In Table 4 are displayed the parameters sets used.

Table 4: Genetic Algorithm Params

Products	Population	Tournament	Generations	Pcr	Pmt
Pr	$Pp = Pr * 10$	$T = \lceil Pp * 0.05 \rceil$	$G = Pp * 5$	0,90	0.07
$Pr = 10$	100	5	500	0,90	0.07
$Pr = 20$	200	10	1000	0,90	0.07

The parameters table follows the order: Quantity of Products, Population Size, Tournament Size (selection method), Maximum number of Generations, Probability of applying Recombination and, finally, Probability of applying Mutation. The first parameter is not specifically the GA, however affects directly the definition of the problem dimension, how bigger the number of products, a greater amount of permutations is possible, and consequently a larger space of search.

SIMULATIONS AND RESULTS

The instance used in the tests was the list of Products $[1, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12]$ and generates the map with the coordinates that correspond to their gondola and slot position, as can be seen in Fig. 2. It is straightforward to see that the points are well represented in the graph, indicating exactly the location of the slot on the map where the product is located.

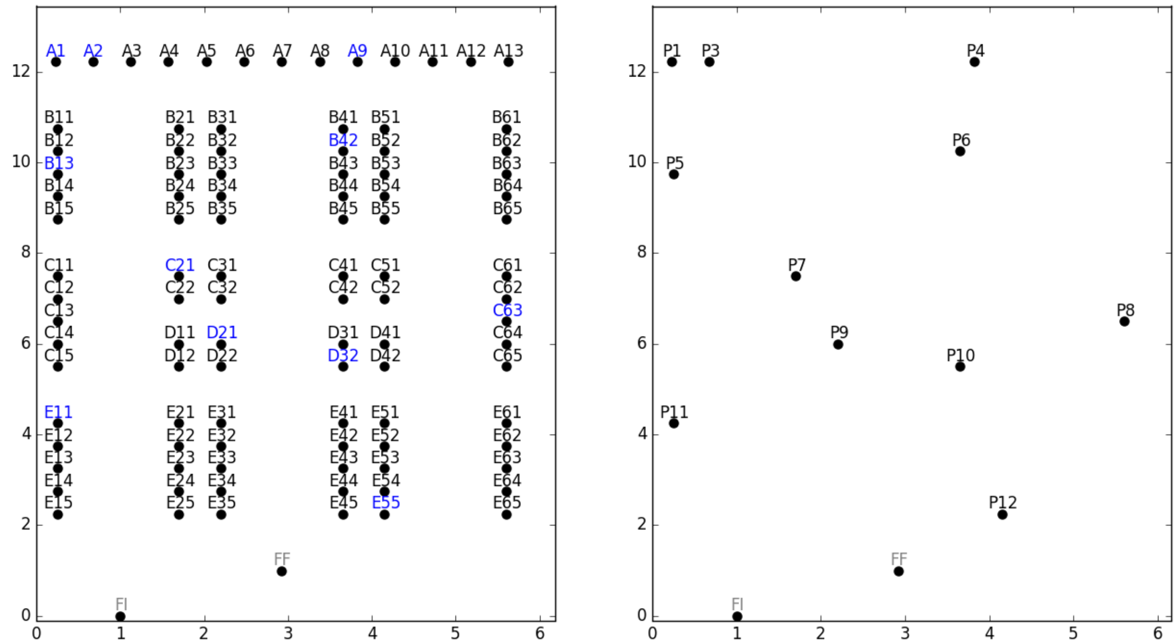

Figure 2: Shelf Layout and Shopping List Product Map

It took an average of 600 generations to bring out the best route, with the fitness value equal 37.53, the permutation $S = [FI, 11, 9, 7, 5, 1, 3, 4, 6, 8, 10, 12, FF]$, where the FI is the initial point and FF is the final point, as can be seen in (Fig. 3).

CONCLUSIONS

In this paper, an adaptive model to solve the SSRPP is proposed, by means of a mathematical model and an evolutionary optimization algorithm. The model is flexible in the sense that the input data are loaded from text files that can be easily generated in the back office system of target retail stores.

Based on these input data the system is able to generate a faithful model of representation. Posteriorly, based on this model, the Genetic Algorithm can solve the SSRPP for each instance (customer shopping list) and the experiments results show that the model can be used to solve the SSRPP successfully.

In future approaches, we intend to include another tools, as beacons, RSSI and geofencing to produce a complete solution that can play a remarkable role in the domain of the location-based indoor services.

ACKNOWLEDGEMENTS

This study was supported by *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior* (CAPES) and *Centro Federal de Educação Tecnológica de Minas Gerais* (CEFET-MG).

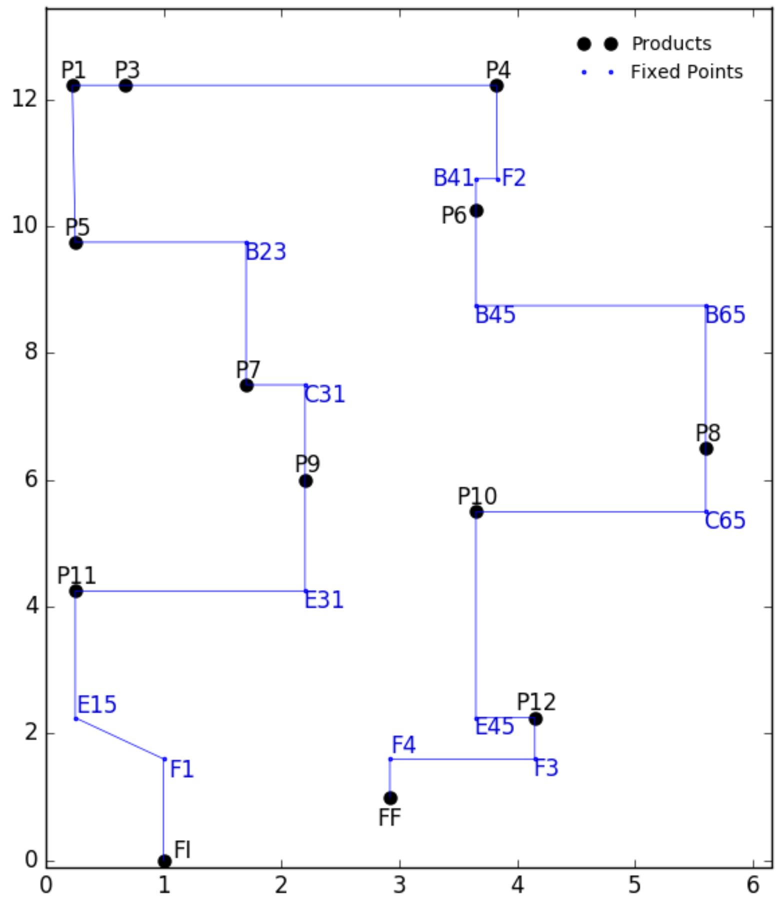

Figure 3: Best route found for the Instance

REFERENCES

Chen, X., Li, Y., & Hu, T., 2015. Solving the supermarket shopping route planning problem based on genetic algorithm. In Computer and Information Science (ICIS), 2015 IEEE/ACIS, *14th International Conference*, pp. 529-533.

Kaneko, Y., Yada, K., 2016. Fractal Dimension of Shopping Path: Influence on Purchase Behavior in a Supermarket. *Procedia Computer Science*, vol. 96, pp. 1764–1771.

Kholod, M., Nakahara, T., Azuma, H., Yada, K., 2010. The Influence of Shopping Path Length on Purchase Behavior in Grocery Store. *Knowledge-Based and Intelligent Information and Engineering Systems: 14th International Conference*, September, pp. 273-280.

Gaspar-Cunha, A., Takahashi, R., Antunes, C. H., 2013. Manual de computação evolutiva e metaheurística. *Imprensa da Universidade de Coimbra / Coimbra University Press*.

Goldberg, D. E., Lingle, R. et al., 1985. Alleles, loci, and the traveling salesman problem. In: Lawrence Erlbaum, Hillsdale, NJ. *Proceedings of an international conference on genetic algorithms and their applications*, vol. 154, pp. 154-159. PMX Crossover Operator.

Syaekhoni, M. A., Lee, C., Kwon, Y. S., 2016. Analyzing customer behavior from shopping path data using operation edit distance. *Applied Intelligence*, Springer, pp. 1-21.