

Space Allocation in the Retail Industry: A Decision Support System Integrating Evolutionary Algorithms and Regression Models

Fábio Pinto and Carlos Soares

INESC TEC/Faculdade de Engenharia, Universidade do Porto*
Rua Dr. Roberto Frias, s/n
Porto, Portugal 4200-465
fhpinto@inescporto.pt, csoares@fe.up.pt

Abstract. One of the hardest resources to manage in retail is space. Retailers need to assign limited store space to a growing number of product categories such that sales and other performance metrics are maximized. Although this seems to be an ideal task for a data mining approach, there is one important barrier: the representativeness of the available data. In fact, changes to the layout of retail stores are infrequent. This means that very few values of the space variable are represented in the data, which makes it hard to generalize. In this paper, we describe a Decision Support System to assist retailers in this task. The system uses an Evolutionary Algorithm to optimize space allocation based on the estimated impact on sales caused by changes in the space assigned to product categories. We assess the quality of the system on a real case study, using different regression algorithms to generate the estimates. The system obtained very good results when compared with the recommendations made by the business experts. We also investigated the effect of the representativeness of the sample on the accuracy of the regression models. We selected a few product categories based on a heuristic assessment of their representativeness. The results indicate that the best regression models were obtained on products for which the sample was not the best. The reason for this unexpected results remains to be explained.

Keywords: Retail, Representativeness of Sample, Evolutionary Algorithms, Regression.

1 Introduction

This paper addresses the problem of assigning space to product categories in retail stores. According to the business specialists, space is one of the most expensive resources in retail [1]. This makes product category space allocation one of the

* Part of this work was carried out while the authors were at Faculdade de Economia, Universidade do Porto.

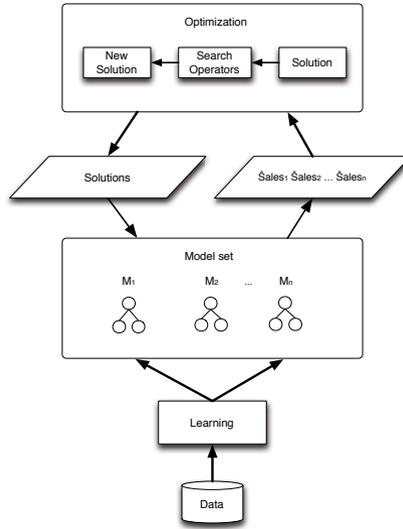


Fig. 1. Decision Support System architecture

most difficult decisions while defining the layout¹ of a retail store. Given that space is a limited resource, it is essential for the retailer to assess the effects of allocating more space to product category x instead of product category y . One approach is based on the estimated space elasticity of product categories [2]. *Space elasticity* is the impact on the sales of a given product or product category of varying (typically 1%) the space allocated to it.

We developed a Decision Support System (DSS) for space allocation that combines machine learning techniques with a meta-heuristic optimization method (Figure 1). The meta-heuristic searches the space of all admissible space allocations. For each solution considered, the corresponding total sales are estimated based on individual forecasts of the sales of each product category obtained using sales forecasting models. Besides the space allocated to the corresponding category in the solution, the inputs to these models include other variables characterizing the store and the product category. The models are induced using machine learning techniques on historical data.

This approach involves a number of challenges, the most important of which are 1) the representativeness of the data, 2) the evaluation of the individual models and 3) the evaluation of the whole system. The data collected by retail companies represents a tiny fraction of its domain because changes to the space assigned to a product category are not frequent. Therefore, it is hard to obtain models with good generalization capacity. To address this issue, we developed

¹ The layout is a schema combining text, graphics and photos to represent the physical distribution of products and product categories in a retail store, as well as their sizes and weights.

a measure of the space volatility of product categories. Surprisingly, however, the best models were not obtained on the product categories with more space volatility according to our measure.

The evaluation of both the individual models and the complete system are also challenging. The development of the individual models is hard because it is difficult to relate their accuracy with the quality of the global store layout. Therefore, knowing when to stop the development process is hard. We addressed this issue by setting thresholds, which were defined in collaboration with the business experts. Finally, the evaluation of the impact on sales of the layouts recommended by the system on real stores is not possible. On the other hand, evaluating them on historical data, even if a suitable resampling methodology is used, is not entirely convincing to the business users. We addressed this issue by using the system to make recommendations for a store that had its layout been recently makeover. We compared our recommendation to the new layout which was implemented, with very satisfactory results according to the experts.

We discuss these issues in this paper, which is organized as follows. Section 2 presents related work. In Section 3 we describe the data used as well as the measure to assess space volatility and the results of its application. Section 4 presents the methodology used to build the sales forecasting models. Sections 5 and 6 describe two experiments in modeling product categories sales: first, only four product categories were modeled, carefully selected according to the space volatility measure; secondly, we model all product categories. In Section 7 we specify how we combined the forecasting models with an optimization algorithm for our DSS. The case study is presented in Section 8, together with the results obtained. Finally, in Section 9 we present some conclusions and define future work.

2 Space Allocation in Retail Stores

Space allocation is done at multiple occasions. Retailers are obviously faced with this problem when opening new stores. Additionally, they must also make seasonal adjustments (e.g., add a camping section before summer) as well as temporary ones (e.g., to accommodate brand promotions). Finally, as the business evolves due to changes in the socio-economic conditions of the customer base (e.g., economic crisis) and new trends (e.g., inclusion of gourmet section), the layout must also be adapted.

Space allocation is done at multiple levels of granularity, namely the shop sections (e.g., vegetables) and all levels of the product hierarchy, including individual brands and products. Depending on the granularity level, decisions may involve location in 3D (i.e., not only the position in the store but also the height of the shelf in which the products are displayed) as well as size (e.g., length of shelf space assigned) and positioning relative to other products. Furthermore decisions are affected by a number of business constraints (i.e., contracts with product manufacturers), which makes the problem of space allocation even more complex.

Thus it comes as not surprise that product space allocation within a retail store is a common research topic. Several studies were conducted for studying the process of product allocation using Econometrics [3], Operations Research [4] and even Machine Learning, with the application of association rules [5] and Genetic Algorithms [6]. However, these papers are not concerned with product category space allocation. They focus on distributing products on previously assigned product category space.

Desmet and Renaudin [2] published the first paper concerning the problem of product category space allocation. They used Econometrics to model sales behaviour of product categories and estimated the respective space elasticity for each. Despite interesting, their results were partially questionable, with some estimated space elasticities with negative value. Castro [1] followed a very similar approach.

We believe that this is the first published work combining optimization and machine learning techniques on this problem.

3 Data

This Section details the data collected for the project and describes the process of identifying the categories with the most representative samples, including the space volatility measure.

3.1 Data Collection

The dataset comprises two years (2009-10) of data with monthly observations for 110 product categories. Overall, the dataset contains 332,885 observations.

The majority of the variables were provided by the retail company in which this project was developed. Due to confidentiality reasons, we can not give insights about their construction but we can motivate the purpose for their inclusion in our dataset. For all variables, i represents a product category, m represents a month and s a store: 1) **Sales** $_{i,m,s}$ is the target variable; 2) **Area_t** $_{i,m,s}$ is the total area,² in square meters, assigned to a product category. [1] showed its significance on sales forecasting models for retail; 3) **Area_pe** $_{i,m,s}$ is the permanent area, in square meters, of a product category. Permanent area does not change due to seasonality factors. This type of area only changes during store layout restructuring; 4) **Area_pr** $_{i,m,s}$ is the promotional area, in square meters, of a product category. Promotional area changes mainly due to seasonality factors; 5) **m** is the month of the example. It is included for seasonality purposes as retail sales are highly seasonal and we expect to model that volatility with this nominal variable; 6) **Insignia** $_s$ is the insignia of the store. The retail company has three different types (insignias) of stores. This nominal variable captures different sales behaviour among these insignias; 7) **Cluster** $_s$ is the sales potential cluster of the store (nominal variable). The retail company divides its stores into

² The total area is the sum of the permanent and promotional areas.

four distinct clusters according to their sales potential; 8) **Cluster_Client_s** is the client profile cluster of the store. The retail company divides its stores in four distinct clusters according to the profile of their customers. Again, the inclusion of this nominal variable seemed relevant given that it is expected that different customers will result in stores with different sales behaviour; 9) **PPI_County_s** is the Purchasing Power Index of the region in which the store is located. It is expected that the larger the value of this variable, the larger the value of sales; 10) **N_W_Days_m** is the number of non-working days of the month. Customers do most of their shopping on non-working days so it is expected that the larger the value of this variable, the larger the value of sales; 11) **C_P_Index_{i,s}** is the category penetration index by store's client profile cluster. This is a discrete variable calculated for each product category within each customer cluster, so, there are 4 indexes by product category, one for each cluster. This variable can capture the impact that different customers have in product category sales.

Although there may have another important factors affecting sales and space, these are the variables that were available for this project.

3.2 Representativeness of Sample

The main goal of the project is to implement an optimization algorithm that maximizes sales given the decision variables **Area_t_{i,m,s}** for all categories i . Therefore, we need models that predict sales accurately over a wide range of values of the latter variables. To achieve this, it is necessary to have data that covers a representative part of the space. However, this is very unlikely, as there are few changes to the space allocated to a category. Given the importance of space to retail, changes must be carefully motivated. The categories that are changed most often may be changed twice a year, while many have constant shelf area over much longer periods than that. Furthermore, most variations are relatively small.

This is an important issue because the quality of the results depends not only on the quality of the variables but also on the representativeness of the training sample. This is illustrated in Figure 2. Samples A and B contain examples in every region of the combined domain of both variables, x_1 and x_2 . However, without further domain knowledge, sample A is better than sample B because it is more dense. Sample C is the worst of the three because only a small subset of the space is represented. The data in our case study is both sparse (as in sample B) and concentrated in a very small area of the domain space (as in sample C).

To understand how serious this problem is in our case study, we developed two measures to characterize shelf space volatility. These measures can be applied to all variables representing space allocation, namely **Area_t_{i,m,s}**, **Area_{pe}_{i,m,s}** and **Area_{pr}_{i,m,s}**, which are represented by $a_{i,m,s}$. The **Mean Number of Changes** (MNC_i) of the shelf space of category i is defined as

$$MNC_i = \frac{\sum_{s=1}^n f_{i,m,s}}{n} \quad (1)$$

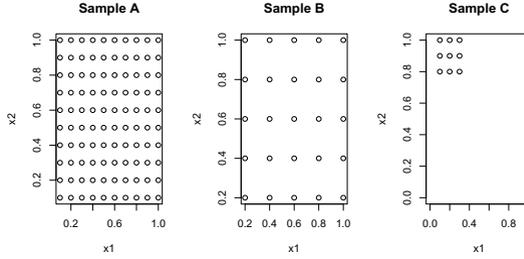


Fig. 2. Sample representativeness

where n is the number of observations and

$$f_{i,m,s} = \begin{cases} 1 & \text{if } a_{i,m,s} \neq a_{i,m-1,s} \\ 0 & \text{otherwise} \end{cases}$$

The **Absolute Mean Monthly Variation** ($AMMV_i$) of the shelf space of category i is defined as (a percent value)

$$AMMV_i = \frac{\sum_{s=1}^n \frac{\sum_{m=2}^k |a_{i,m,s} - a_{i,m-1,s}|}{k}}{n} \times 100 \tag{2}$$

The NMC_i and $AMMV_i$ measures are combined into a space volatility score, defined as $SV_i = MNC_i \times AMMV_i$.

Figure 3 shows the values of MNC_i and $AMMV_i$ for each product category in our dataset. The categories with a score value equal or greater than the 90th percentile are considered to have a high score level and are represented in red. As hypothesized, a great part of the product categories lie in the lower left side of the chart, meaning that those categories have reduced space volatility.

However, a high volatility score is not sufficient to ensure that the data sample for the corresponding category is representative of its domain. It is also important that the category has a homogeneous sales behavior across stores. The purpose of this requirement is to reduce the impact of the factor *store* in the relationship between product category space and sales. To do this, we compute the standard deviation of sales as a percentage of total sales by product category. The lower this value, the more homogeneous are the sales of the product category in the set of stores.

The graph of Figure 4 illustrates different scenarios that were found. It presents the indicators for the eleven product categories with a high volatility score: on the x-axis, the score for the permanent area of the categories; on the y-axis, the score for the promotional area; and the size of each point relates to the store homogeneity value of the respective product category, which are distinguished by different colours.

Interesting categories are: **3204**, with high volatility in terms of permanent area; above-average store homogeneity; **3101** with high volatility in terms of

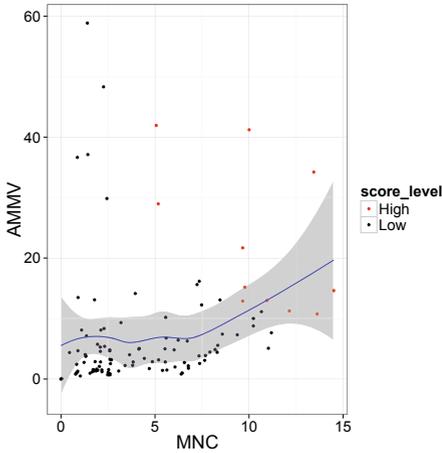


Fig. 3. Measures representing MNC_i and $AMMV_i$ for each category i . The categories with a high value of the score SV_i are represented in red color.

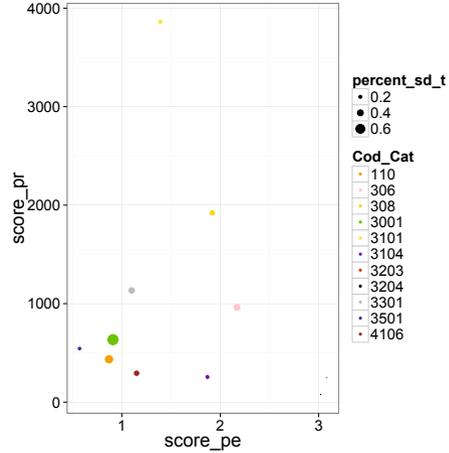


Fig. 4. Store homogeneity (size of points) for the categories with high sales volatility (SV_i) in permanent (y -axis) and promotional (x -axis) areas

promotional area and store homogeneity slightly below average; **3001** with considerable volatility in both area components and which is the product category with the best record of store homogeneity among the categories with a high volatility score; and **308**, with considerable volatility in both area components and which it is the product category with the worst record of store homogeneity among the selected categories. The latter category is the one with the worst data sample, according to our measures, while the other have the best samples.

4 Methodology

The problem is addressed as a regression task with $Sales_{i,m,s}$ as the dependent variable. Next, we describe how error is estimated and which regression algorithms were used.

4.1 Error Estimation

The error measures for the evaluation of the predictive models are: *Mean Relative Error (MRE)*, defined as $\frac{\sum_{j=1}^n \frac{|y_j - \hat{y}_j|}{y_j}}{n} \times 100$; *Root Mean Squared Error (RMSE)*, defined as $\sqrt{\frac{\sum_{j=1}^n (y_j - \hat{y}_j)^2}{n}}$; and *Variation Index (varIndex)*, defined as $\frac{RMSE}{\bar{y}}$. For all measures, y_j is the true value, \hat{y}_j is the predicted value, and \bar{y} is the mean of the output variable.

The performance of the regression algorithms will be compared with two benchmarks: a linear regression (LR), given that this technique was applied

in past experiments on this problem [1][2]; and a baseline, whose predictions consist of the average of the target variable for the store which is being analyzed. These comparisons will help assess the difficulty of the phenomenon that we are modeling as well as how much useful knowledge the regression models are capturing.

The retail company defined 10% as the (maximum) target MRE value. We used this value as a threshold for success of our models.

Given that the dataset used in this work consisted of time series, the error was estimated using a hold-out strategy, ensuring that the test data was more recent than training data. The training set for each product category (and each model) consisted of one year and four months of observations; the remaining examples (eight months) were split (4 months each) for a validation and a test set. The validation set was used for algorithm parameter tuning and the test set to assess the generalization error of the models.

4.2 Regression Algorithms

Several regression models were tested, namely: **Cubist**, an improvement of Quinlan's M5 regression trees [7]; **Artificial Neural Networks** (ANN), based on the empirical evidence of the capacity of ANN to successfully predict retail sales [8]; **Multivariate Adaptive Regression Splines** (MARS) [9]; **Support Vector Machines** (SVM) [10]; **Generalized Boosted Models** (GBM), *R* package implementation of boosting [11] models; and **Random Forests** (RF) [12].

The implementation of these algorithms available in the *R* [13] software was used in the experiments.

5 Modeling Categories with High Sales Volatility

The measures that were proposed in Section 3.2 to quantify space volatility were validated by the business experts. Nevertheless, we decided to analyze empirically if they are, in fact, good measures of the representativeness of the data samples. In case of positive results, we can define a strategy to systematically collect data for the product categories with data of insufficient quality based on the space volatility measures.

We focused on the four categories that were identified in Section 3.2, namely, **3204**, **3101**, **3001** and **308**. The results obtained on the category **308** are expected to be worse than on the others, given that the computed measures of space volatility and store homogeneity are clearly worst for this product category. Different parameters settings were tested and the best results are presented here.

Figure 5 shows the results for the four selected product categories in terms of MRE. Surprisingly, the MRE estimated for the models is far from the target value of 10%, except for product category **308**, which is actually quite close to the threshold. These results do not confirm our hypothesis. There seems to be no correlation between the variables calculated to characterize the representativeness of the data for each product category and the predictive performance

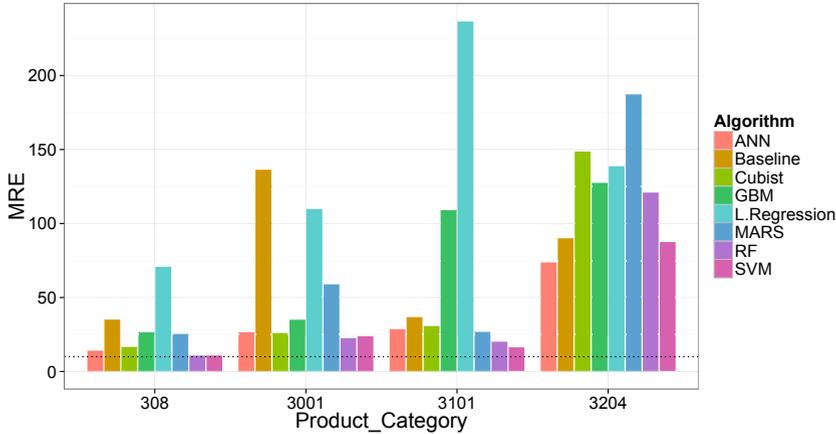


Fig. 5. Results for selected product categories, in terms of MRE. The dotted horizontal line represents the threshold defined by the client.

of the algorithms. Further analysis provided possible explanations for some of these results. For instance, the product category with the worst results is **3204**. Given that the models of this category do not include the variable m , since there was only one year of data for this category, we assume that the problem may be lack of representative data, as retail sales are seasonal. Nevertheless, at this point it became clear that we could not yet develop the data collection strategy based on the current version of these measures.

Table 1 presents more detailed results, according to three error measures. In terms of performance of the regression algorithms, SVM, RF and ANN obtain the best results. Overall, for these product categories, SVM is the best algorithm, according to the average of the error measures. Further observations can be made: 1) for all product categories, there is always at least one regression algorithm that is better than both benchmarks; 2) ANN presents reasonable performance from modeling retail sales, confirming previous results [8]; and 3) LR obtains poor performance for modeling retail sales. This fact is particularly important if we remember that the previous papers on the topic of product category space allocation [1][2] used this technique.

6 Modeling All Product Categories

Given the unexpected results in the preliminary experiments, we decided to run experiments on all product categories. Given that some of the 110 product categories of our dataset showed a very low number of observations, we only used the 89 categories with more than 1000 observations. Additionally, for 10 of these product categories, we only had one year of data.

We decided not to test ANN because of its computational cost. Additionally, the best SVM results in the previous section were obtained using two kernels,

Table 1. Detailed results on selected product categories results

Algorithm	Error Measure	Product Categories				Average
		3204	3101	3001	308	
Cubist	MRE	149.02%	31.10%	26.64%	16.97%	55.93%
	RMSE	300.57	4160.69	22118.84	6218.20	8199.58
	varIndex	182.33%	48.05%	44.12%	28.96%	75.86%
ANN	MRE	74.09%	28.97%	26.96%	14.49%	36.13%
	RMSE	208.57	2238.49	26816.24	5809.25	8768.14
	varIndex	135.97%	25.08%	60.33%	39.27%	65.16%
MARS	MRE	187.64%	27.16%	59.30%	25.70%	74.95%
	RMSE	348.37	2437.43	31767.91	7055.82	10402.38
	varIndex	211.33%	28.15%	63.36%	32.86%	83.93%
SVM	MRE	87.84%	16.70%	24.19%	11.22%	34.98%
	RMSE	94.81	2153.42	18680.24	5458.18	6596.66
	varIndex	57.52%	24.87%	37.26%	25.42%	36.27%
GBM	MRE	127.86%	109.4%	35.36%	26.93%	74.89%
	RMSE	220.68	6414.26	30985.78	15500	13280.18
	varIndex	127.78%	74.50%	59.28%	64.93%	81.62%
RF	MRE	121.28%	20.53%	22.92%	11.12%	43.96%
	RMSE	151.23	3044.51	19620.52	8373.84	7797.53
	varIndex	90.26%	35.20%	38.36%	38.28%	50.53%
LR	MRE	139.10%	236.87%	110.20%	71.13%	139.33%
	RMSE	264.26	5299.86	30761.88	18619.4	13736.35
	varIndex	160.31%	61.20%	61.35%	86.72%	92.40%
Baseline	MRE	90.40%	37.14%	136.72%	35.45%	74.93%
	RMSE	136.22	6485.44	78004.39	26796.21	27855.56
	varIndex	82.64%	74.90%	155.56%	124.80%	109.48%

therefore, we decided to use those two different kernels. So, we tested three algorithms: SVM with the radial and sigmoid kernels and RF.

6.1 Results

Figure 6 shows the results obtained for the three regression algorithms in terms of MRE on 89 product categories. The dotted horizontal line represents the threshold defined in Section 4. Surprisingly, given the results obtained earlier, several models are below or very close to the threshold. On one hand, this confirms that the measures used to assess space volatility need to be improved but, on the other, it indicates that the approach followed is viable, regardless of the apparent lack of representativeness of the data.

The statistical significance of the differences was based on the results of paired t-tests with $\alpha = 0.05$. Overall, SVM with the radial kernel obtained the best performance in 77 product categories; RF obtained the best performance in 10 product categories; and finally, SVM with sigmoid kernel presented superior

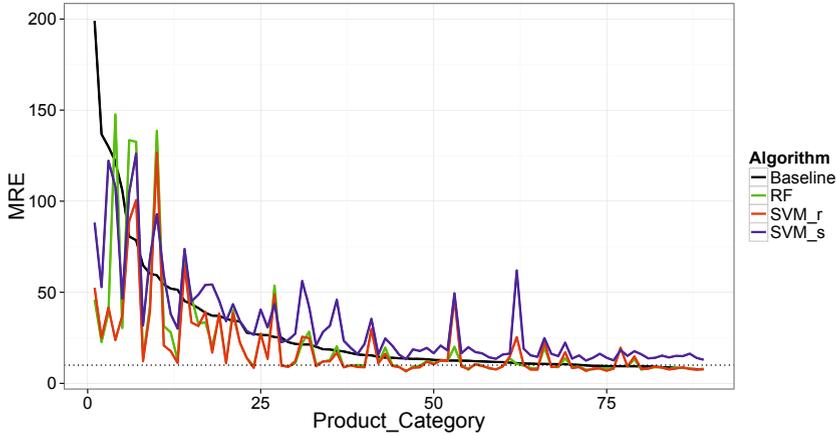


Fig. 6. Results of SVM - radial kernel, SVM - sigmoid kernel and RF for all product categories

performance in only two product categories. In several product categories, RF and SVM with radial kernel function showed a very similar performance and the statistical test did not distinguish them.

In Table 2 we compare the best model for each product category (of the three that were generated) with the respective baseline. On average, the regression algorithms show a better predictive performance than the baseline. More importantly, approximately half of them obtained an error that was very close or better than the threshold set by the retail company.

Table 2. Summary: best model vs Baseline (MRE)

Algorithm	Min	1stQ	Median	Mean	3rdQ	Max
Best model	6.51	8.76	11.01	19.05	22.95	126.11
Baseline	7.69	10.58	13.78	27.73	27.83	198.90

7 Decision Support System

Given that the models presented results that were very close to the goal set by the business experts, we decided to develop the method to optimize the assignment of shelf space to the product categories. As explained in more detail next, this method uses the models to predict the sales for the shelf space assigned by the solutions tested in the search process.

The optimization algorithm is an adapted version of a Genetic Algorithm (GA), an Evolutionary Computation (EC) framework. For further details on GAs, we refer the reader to [14].

Algorithm 1 shows the pseudocode of the algorithm implemented. The general EC method must be adapted to take into account specific issues of the problem at hand. In our case, solutions must not violate three fundamental constraints: a minimum value for each gene (the minimum space of a certain product category); a maximum value for each gene (the maximum space of a certain product category); and a maximum value for the sum of all genes (the maximum space of the sum of all product categories and layout).

```

begin
  INITIALISE population with historical and random solutions;
  EVALUATE each candidate;
  repeat
    SELECT parents;
    RECOMBINE pairs of parents with whole arithmetic recombination;
    MUTATE the resulting offspring;
    EVALUATE new candidates;
    SELECT top 30 individuals for the next generation;
  until number of iterations = it;
end

```

Algorithm 1. Optimization algorithm.

Representation of the solutions. A solution consists of the area of all product categories in a given store, $sol_h = g_{1,h}, g_{2,h}, \dots, g_{i,h}$. Each solution has as many elements as the number of product categories.

The initial population consists of 30 solutions: 10 are historical assignments of shelf space to product categories in the store that is being optimized; and the remaining 20 are random solutions. The range of valid space values for each product category is defined by the corresponding maximum and minimum values that occurred in the most recent year available in the dataset for the store that is being optimized: $min_i \leq g_{i,h} \leq max_i$.

The fitness function that evaluates the solutions is based on the predictive models generated in Section 6. It uses those models to estimate the sales of the product categories, given the shelf size in the solutions together with the values of the remaining independent variables. The fitness of a solution is the sum of all sales forecasts, plus a penalty function ω_h for controlling the total space of the solution: $Fitness_h = sales_1 + sales_2 + \dots + sales_i + \omega_h$.

Parents selection. The selection of the solutions that will generate *offspring* is based on a probability associated with each, taking into account the fitness function output for each solution: $Prob(k) = Fitness(k) / \left(\sum_{k=1}^{30} Fitness \right)$. Then, from 30 solutions that constitute the population, 10 are selected for crossover.

Crossover. In this stage, the 10 selected solutions, generate 10 new solutions. In order not to violate the constraints that were imposed on all solutions, we

applied an operator named whole arithmetic crossover [14]: given two selected solutions, sol_1 and sol_2 , the average of the two is calculated, originating a new solution: $sol_3 = \frac{g_{1,1}+g_{1,2}}{2}, \frac{g_{2,1}+g_{2,2}}{2}, \dots, \frac{g_{1,h}+g_{1,h}}{2}$.

Mutation. This operator randomly selects one or more solutions from the offspring. Two genes are randomly chosen from each of the selected solutions, and 1 unit (square meter) is transferred from the first category to the second: $sol_3 = g_{1,3} + 1, g_{2,3} - 1, \dots, g_{h,3}$. However, this operator may disregard restrictions on the minimum and maximum area of each product category. In order to circumvent the problem, the mutation only occurs if the selected gene does not have a minimum or maximum value. The number of offspring selected for mutation is controlled by the mutation rate parameter.

Survival Selection. After the operators are applied, the offspring are added to the population. With a total population of 40 solutions, a new evaluation by the fitness function occurs. The top 30 are kept for new iteration. In this work we have opted for an elitist approach in the survivor selection mechanism [14], instead of the typical probabilistic one. We chose this approach given that the fitness values for our solutions were very close to each other, and a probabilistic selection lead the algorithm to the loss of good solutions.

The DSS performs optimization at the monthly level: given a store, a particular number of product categories and one month, the algorithm seeks the set of areas that maximizes monthly sales according to the predictive models generated in Section 6.

8 Case Study

The retail company proposed a test for a more realistic assessment of the system. In May 2011, a given store had a makeover of its logistics which included a new layout. The vast majority of the product categories that were available in this store underwent major changes in the shelf space. For this makeover, the analysts of the retail company based their space recommendations on data from 2009 to 2010, the very same that allowed us to build our dataset. Thus, it was proposed to test the developed DSS in this store and compare the results obtained with the recommendations of the business specialists.

8.1 Experimental Setup

In this experiment, we assumed that, except for the variables $Area_{t_i,m,s}$ and $N_W_Days_m$, the other independent variables remained constant, given that it seems acceptable that from 2010 to 2011 they have not changed substantially. The predictive models generated for this experiment integrated all observations available in the dataset. Based on the experiments in Section 6, we selected the best algorithm for each product category.

Out of the 89 categories for which it is possible to build predictive models, 85 are part of the store layout, so, the solutions consisted of 85 genes. To evaluate the quality of the optimization and compare its results with recommendations from the business specialists, we used Pearson's correlation coefficient.

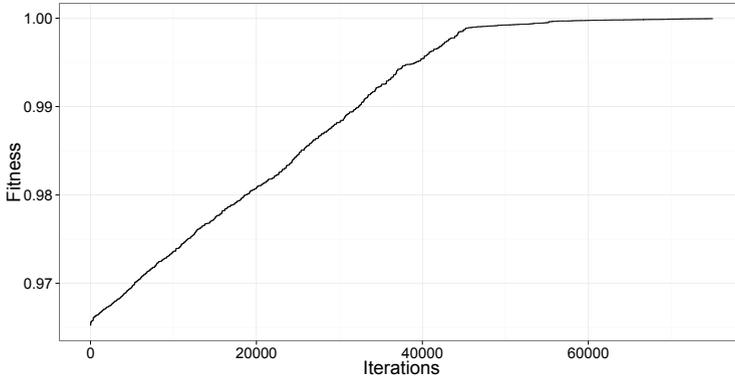


Fig. 7. DSS results for 75000 iterations, measured as the fitness of the best solution evaluated

8.2 Results

For 75,000 iterations, the optimization algorithm presents the performance showed in Figure 7. Due to confidentiality issues, the fitness values were scaled between 0 and 1, 0 being the minimum fitness value and 1 being the maximum fitness achieved by the best solution in the population. There is a steep increase in fitness, which slows down as the system converges. A steady state is achieved near iteration 60,000.

Comparing the results obtained by the DSS with the recommendations made by the business specialists, we estimated a correlation of 0.66. Assuming that the space recommendations of the business specialists are ideal, these results are a good indicator of the quality of the developed DSS. We compared the predicted total sales according to our models of the two solutions.

Analysing the results in detail, some important questions arise. A first comparison between the output of the DSS and the real areas of the store in May 2011, shows that three product categories are clearly overvalued by the system. These categories have a common factor: they have very strong promotional campaigns at specific times of the year. It is clear that the models of these product categories (with MRE values of 11.17%, 24.90% and 16.80%) are failing to associate this boom in sales to seasonal factors. Those strong promotional campaigns also imply a significant, albeit temporary, increase in shelf space, which seems to be confusing the learning algorithms. Given that the m variable (month of the observation) varies systematically and the variable $Area_{t_i,m,s}$ presents a large increase at those times just like the variable $Sales_{i,m,s}$, the forecasting models associated the sales boom to changes in the value of the variable $Area_{t_i,m,s}$. This has implications in the DSS output: for these product categories, its recommendation is to have promotional campaigns at any time of the year. This clearly does not make sense in seasonal products.

Although, we assume that the recommendations made by the specialists are ideal, due to lack of better information, the product categories for which the system differs from those values are also interesting. For instance, for the adjacent product categories 1502 and 1503, the DSS suggests areas that are quite different from the recommendations of the business specialists. Given the excellent predictive accuracy of the models in question (6.92% and 12.17%, respectively), we found this information to be quite useful for the retail company.

9 Conclusions and Future Work

This paper presents the combination of sales forecasting models for individual product categories and a GA to develop a DSS for product category space allocation in retail stores. As far we know, this is the first time that machine learning and optimization techniques are applied to deal with this problem.

Given that not many changes are made to the layout of stores, it is expected that the data collected from the daily operation of retail stores is not adequate for modeling purposes. We developed two measures to assess the representativeness of the data associated with each product category. However, an empirical study indicated that the measures were not good predictors of the predictive accuracy of the models. Given the importance of collecting adequate data, an important goal is to improve these measures to serve as the basis for a systematic data collection strategy. This work can benefit from existing work on active learning [15].

Somewhat surprisingly, many of the models generated obtained satisfactory results and, thus, we developed the DSS. The system was evaluated with data from a major change in layout at a specific store, which had recently been performed. The results indicate that this system recommends space allocations in retail stores that are very similar to the recommendations made by business specialists. Furthermore, some of the differences in the recommendations were interesting from the business perspective.

We have developed models that are independent of each other. However, it is likely that there is some dependence between the sales of at least some of the product categories. We plan to test multi-target regression methods [16] to address this problem.

In this project, we have addressed only a very small part of the space allocation problem. An important development is the extension of the current system with recommendations concerning which categories to place next to each other. One approach to this problem is market basket analysis [17]. However, the challenge is how to combine the two types of recommendation.

Acknowledgements. This work was partially supported by Project Best-Case, which is co-financed by the North Portugal Regional Operational Programme (ON.2 - O Novo Norte), under the National Strategic Reference Framework (NSRF), through the European Regional Development Fund (ERDF) and by

National Funds through the FCT - Fundação para a Ciência e Tecnologia (Portuguese Foundation for Science and Technology) within project “Evolutionary algorithms for Decision Problems in Management Science” (PTDC/EGEGES/099741/2008).

References

1. Castro, A., Brochado, A., Martins, F.: Supermarkets sales and retail area: a mixture regression model for segmentation. In: *European Network for Business and Industrial Statistics* (2007)
2. Desmet, P., Renaudin, V.: Estimation of product category sales responsiveness to allocated shelf space. *International Journal of Research in Marketing* (15), 443–457 (1998)
3. Gaur, V., Fisher, M., Raman, A.: An econometric analysis of inventory turnover performance in retail stores. *Management Science* 51, 181–193 (2005)
4. Dréze, X., Hoch, S.J., Purk, M.E.: Shelf management and space elasticity. *Journal of Retailing* 70(4), 301–326 (1994)
5. Nafari, M., Shahrabi, J.: A temporal data mining approach for shelf-space allocation with consideration of product price. *Expert Systems with Applications* (37), 4066–4072 (2010)
6. Hwang, H., Choi, B., Lee, G.: A genetic algorithm approach to an integrated problem of shelf space design and item allocation. *Computers and Industrial Engineering* (56), 809–820 (2009)
7. Quinlan, J.: Learning with continuous classes. In: Adams, Sterling (eds.) *AI 1992*, Singapore, pp. 343–348 (1992)
8. Alon, I., Qi, M., Sadowski, R.J.: Forecasting aggregate retail sales: a comparison of artificial neural networks and traditional methods. *Journal of Retailing and Consumer Services* 8(3), 147–156 (2001)
9. Friedman, J.: Multivariate adaptive regression splines. *The Annals of Statistics* 19(1), 1–141 (1991)
10. Vapnik, V., Cortes, C.: Support-vector networks. *Machine Learning* (20), 273–297 (1995)
11. Friedman, J.: Greedy function approximation: a gradient boosting machine. *The Annals of Statistics* 29(5), 1189–1232 (2001)
12. Breiman, L.: Random forests. *Machine Learning* 45(1), 5–32 (2001)
13. R Core Team: *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria (2012)
14. Eiben, A., Simth, J.: *Introduction to Evolutionary Computing*, 1st edn. *Natural Computing Series*. Springer (2003)
15. Settles, B.: *Active learning literature survey*. Technical report, University of Wisconsin, Madison (2010)
16. Aho, T., Zenko, B., Dzeroski, S., Elomaa, T.: Multi-target regression with rule ensembles. *Journal of Machine Learning Research* 1, 1–48 (2012)
17. Russell, G.J., Petersen, A.: Analysis of cross category dependence in market basket selection. *Journal of Retailing* 76(3), 367–392 (2000)