Adaptive Business Intelligence in healthcare - A platform for optimising surgeries

José Ferreira[1], Filipe Portela[1], José Machado[2] and Manuel Filipe Santos[1]

[1]University of Minho, Department of Information Systems
[2]University of Minho, Department of Informatics

Abstract. Adaptive Business Intelligence (ABI) combines predictive with prospective analytics in order to give support to the decision making process. Surgery scheduling in hospital operating rooms is a high complex task due to huge volume of surgeries and the variety of combinations and constraints. This type of activity is critical and is often associated to constant delays and significant rescheduling. The main task of this work is to provide an ABI based platform capable of estimating the time of the surgeries and then optimising the scheduling (minimizing the waste of resources). Combining operational data with analytical tools this platform is able to present complex and competitive information to streamline surgery scheduling. A case study was explored using data from a portuguese hospital. The best achieved relative absolute error attained was 6.22%. The paper also shows that the approach can be used in more general applications.

Keywords: Decision Support Systems, Adaptive Business Intelligence

Introduction

The idea that moves this work is the absence of Adaptive Business Intelligence (ABI) approaches in hospitals and other health institutions to improve the quality of service through efficient scheduling of surgeries in the decision process. Nowadays, decision support systems integrate prediction and optimization capabilities making it possible to minimise or maximise a specific goal. Complex processes to be solved by humans, can be addressed by computer based decision support systems, in a more straightforwardly way, allowing for a better efficiency and effectiveness of the decisions, saving human and financial resources. ABI systems are decision support systems, capable of predicting, optimising and adapting to external changes.

Work has been conducted in order to extend the capabilities of an already existing hospital business intelligence platform with ABI features. Real data collected from the hospital were used as a proof of concept of the platform developed. The hospital will benefit from an improvement in the quality of the information they have. The platform can help predicting the surgery time and optimisating surgeries scheduling along the shifts vailable. This solution allows the minimisation of time wastes of the total duration of the necessary shifts and avoid the delays that are generated by the lack of use of shifts. This paper is divided into six sections, beginning with an introductory section and fininshing with a section dedicated to conclusions. The related concepts and existing ABI solutions are presented in second and third sections respectively. The case study and their results are detailed are detailed in the remaining sections.

Background

2.1 Adaptive Business Intelligence and Data Mining

Adaptive Business Intelligence is the combination of a business intelligence system, data mining, prediction methods and optimisation techniques, that is, it combines prediction, optimisation and adaptability into a system capable of answering two fundamental questions [1]: What is likely to happen in the future? And what is the best decision right now? This system is used to solve many business problems in the real world, ranging from demand prediction, scheduling, fraud detection and investment strategies to significant benefits and savings [2].

Adaptive Business Intelligence brings together the techniques and tools of the database, data warehouse, data mining, prediction, optimisation and adaptability to increase versatility. Furthermore, it allows business managers to make better decisions, thus improving efficiency, productivity and competitiveness. The Adaptive Business Intelligence System is divided into three components [2]: 1. Prediction (e.g. projections of the standard time of the event);2. Decision making - almost perfect (e.g. scheduling of surgeries) and 3. Adapting predictions and optimisation of external changes. The optimisation is a technique of searching for better parameter values. The purpose of this technique is to find the parameter values that minimise the prediction error, based on the prediction model data. This technique maximizes the number and duration of surgeries allocated to the available shifts based on the predicted duration of the surgeries [2].

Data mining is a science of exploring large amounts of data, which aims to find consistent standards. With Data Mining it is possible to extract implicit and unknown information before exploitation, making it useful to solve problems of various applications in domains like business, health, science and engineering [4, 7]. Daryl Pregibons described the concept of data mining as a " blend of statistics, artificial intelligence, and database research" [3], is still currently the affirmation. The Exploratory Data Analysis (EDA) techniques used in data mining are divided into two parts, which are [4-7]: Computational Methods (e.g. statistical, classification, regression, others) and Data Visualisation.

2.2 Modern Optimisation

Modern optimisation is the name given to the methods known as meta-heuristics, and, as mentioned above, they are applied to minimise or maximise a solution to obtain a satisfactory result that solves a problem. The problems that modern optimisation proposes to address are complex and do not have any specialized optimisation algorithm. They include problems with discontinuities, dynamic changes, multiple objectives or hard and soft constraints that are difficult to manipulate [8] is because the hard constraint cannot be violated because of factors involving laws and physical, and soft constraint can only be adjusted by multi-objective optimisation [8], so that optimisation is always useful in several business areas. During the implementation of a modern optimisation method the user should ever consider the following aspects [2]: representation of the solution; objective function; evaluation function.

Scheduling Optimization Platform

3.1 Method and Tools

The platform was built using Weka tool for the prediction block and RStudio tool for the optimisation block. The Weka tool is an open source software issued under the GNU General Public License, which aims to add algorithms for machine learning for data mining tasks, which brings together various techniques, particularly for this case is the regression being through this technique that is done to predict the duration (default time) of the event. RStudio is a more user-friendly tool for the R language, which allows the use of modern optimisation methods as a solution to the problem, with the purpose of local search optimisation being hill climbing and simulated annealing.

3.2 ABI Architecture

This platform was developed based on the Adaptive Business Intelligence architecture. This platform consists of two blocks, namely a prediction block, which predicts the duration of each surgery or more generally an event, and an optimisation block, which aims to minimise waste of shifts, preventing delays and increasing efficiency and effectiveness of schedule. Figure 1 shows the full flow of the scheduling optimisation platform wherein the next subchapters (3.3 and 3.4) are explained in detail.

3.3 Predicting Block

The first step is to predict the standard time of each event (surgery), through the training of models in data mining, in the Weka tool. In the ABI platform prediction block are used algorithms for machine learning regression, predictive modelling in Weka tool. The techniques used are as follows: Linear Regression (LinearRegression); k-Nearest Neighbours (IBk); Decision Tree (REPTree); Support Vector Regression (SMOreg); Multi-Layer Perceptron (MultilayerPerceptron).

Before training these models, it is necessary to identify which variables are more correlated to the target variable. The target corresponds to the event that is intended to be predicted, in this case the duration of the surgeries. Feature selection is obtained through the algorithm "CorrelationAttributeEval", which makes a ranking of the occurrences of those, and then used the 10 to 15 most relevant characteristics. After this phase, the estimators are trained, to determine which one of them has a lower error rate.

After choosing the model with the lowest error, Weka tool was used to fill the surgery duration for the data set that will be used for optimization.

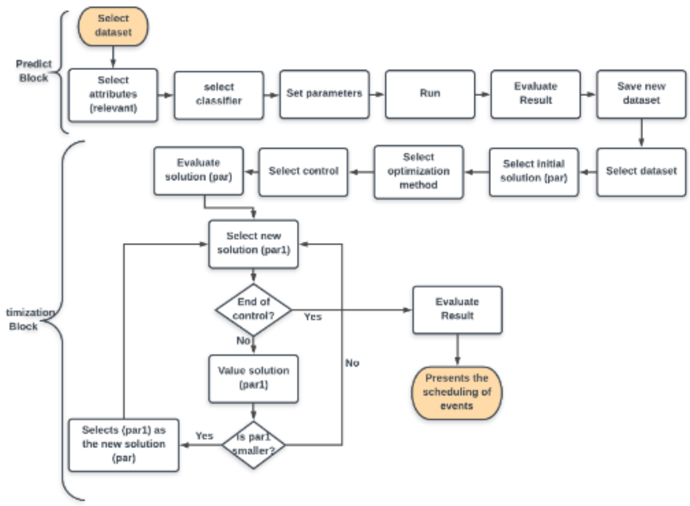

Figure 1 - Platform Architecture Flow

3.4 Optimisation Block

The result of the prediction block is a new dataset containing events with the duration time predicted in the previous block. The methods used in this block are hill climbing and simulated annealing, which require the use of a script file containing the functions called "hclimbing" and "optim" to obtain the best result, which in this case is the less waste between the two methods. The functions of the algorithms and the input parameters for hill climbing method and the simulated annealing method were set as [8]: hclimbing (par, fn, change, lower, upper, control, type) and optim (par, fn, method, gr, control). The initial solution (par) was obtained randomly or sequentially shifts through the list of surgeries of specialities matching the shift, assigning each dimension index a turn. The lower and upper parameters represent the lowest and highest values for each dimension, varying from 1 to the number of shifts of a dimension. The evaluation function calculates the wastes of each shift by subtracting the total standard time from the assigned shift and returns the total sum of shift wastage. It is also in this part of the function that the constraints are defined and the unwanted paths are penalized by assigning a high value to the waste so that this solution is not chosen. The constraints found for the proper functioning of the optimisation were: Date of the event < Date of shift and Event deadline > Date of shift and Waste >= 0

The method is consists of a finite number of iterations through the control function, which contains a list of braking and controls the maximum number of iterations and the frequency control information. The simulated annealing method has only the difference on the trace parameter (set to true), which displays the trace information about the optimisation progress only if they are positive. The iteration limit will be the number that will go through the optimisation and the control and allows for the presentation of a report containing the solution of each one of the iterations, where each following solution is changed through the change function.

The change function aims to generate the next solution, which will undergo minor changes to the previously generated solution, and so on. This function has the following input parameters [8]: hchange (par, lower, upper, dist, and round).

To obtain the optimum value in the hill climbing method is used the minimization option since the goal is to minimize the wastage of the shifts of each surgical speciality. In order to understand the best model to be used as an optimisation method, it will be necessary to compare the two solutions obtained, choosing what for each surgical speciality presents the smallest waste, avoiding delays in the surgeries and, consequently, expenses in unnecessary shifts.

Case Study

A case study will be presented making use of CRISP-DM methodology [9] and a dataset of a Portuguese health institution.

4.1 Business understanding and

The first phase of CRISP-DM is a kind of problem statment. Previous works in this area [9,10] only adressed the time duration prediction. Did not address the optimisation problem and did not solve the wasting of time of the shifts. This can result in a great inefficiency. In this work, the immediate goal is to minimise the waste of each shift, for each one of the surgical specialities, using only the necessary shifts. The estimation of the value for the duration of the surgical specialities is used in order to proceed with the optimisation based on this standard time. For this, regression machine learning algorithms were used to predict the standard time of each surgery, and the satisfactory solution will be achieved using a modern optimisation method. This method of optimisation finds the solution through the neighbouring settlements, that is, it goes through several iterations, calculating the waste of each shift until finding the smallest value for this problem.

This way, it is avoided that future surgeries are delayed due to the unavailability of operating rooms and shifts, since this platform, through the study of the various possible combinations, will select the most satisfactory solution, in the sense of maximising efficiency and effectiveness.

4.2 Data Understanding and Preparation

Due to data set limitations, the tests of each block were performed with different datasets. A dataset was used to predict the duration time of the surgeries and a different dataset was used for the optimisation . This division happens because the table of surgeries available to being used in the optimisation block does not have sufficient surgeries for the forecasting process. Nevertheless, and despite the data used for the tests being different, the cohesion of the platform is not impaired, since the end user has enough data for the forecast result to be used as the standard time for the surgeries in the optimisation. Table 1 presents some statistics of the dataset.

Table 1 - Sample of the statistical analysis of some specialities (time in seconds)

Speciality	Number of Shifts	Quantity of Surgeries	Standard Time Shifts	Standard Time Surgeries	Total Waste
OPHTHALMOLOGY	1078	3162	34905600	5307459	29598141
PEDIATRIC OPHTHALMOLOGY	23	30	909900	116817,8	793082,2
VASCULAR SURGERY	569	796	19389000	3456938	15932062
ORTHOPEDICS	875	1440	26650800	9509328	17141472
PEDIATRIC ORTHOPEDICS	94	163	3187800	848623,6	2339176,4

4.3 Modelling

Regression models (REPTree, LinearRegression, IBk, MultilayerPerceptron and SMOreg) were induced using the Weka tool to predict the operating time in the operating room. At this stage, for testing the prediction block, was used data from the surgical specialities that included the most significant number of surgeries in the entire dataset, namely the orthopaedic speciality. Tables 2 and 3 identify the variables that appear to have a more descriptive power by the occurrence of only a few values, selecting the 15 attributes most relevant for the predicting task.

Table 2 - Attributes with more descriptive power

Attributes	Punctuation
TEMPO6	0.969341
TEMPO7	0.942850
TEMPO2	0.677621
TEMPO8	0.376831
SERVICE	0.345452
GDH	0.299350
MCONBASE	0.296299
COD_ANESTESIA	0.233695
PROCS1	0.233320
INTERVENCAO1	0.232584
COD_INTERV_CIRURGICA	0.230257
COD_SALA	0.144060
PROCS8	0.117009
COD_PATOLOGIA	0.092911

Table 3 - Summary of the regression with 15 attributes

REPTree	IBk	SMOreg	LinearRegression	MultilayerPerceptron
Correlation coefficient	0.9958	0.813	0.9972	0.9955
Mean absolute error	269.7897	2107.5949	214.699	313.8986
Root mean squared error	405.5077	2850.6335	332.11	425.2081
Relative absolute error	7.8189 %	61.0813 %	6.2223 %	9.0973 %
Root relative squared error	9.2617 %	65.1079 %	7.5853 %	9.7117 %
Total number of instances	237	237	237	237

For the optimisation task, modern optimisation methods were implemented, as mentioned above, to obtain the minimum value of the total wastes of the shifts:

- hclimbing (par, fn, change, lower, upper, control, type);
- optim (par fn, method, gr, control).

These methods were used for 32 surgical specialities, allowing annual and monthly optimisation, to find the best solution. The first step is to obtain the initial solution, given randomly or sequentially, through the necessary shifts and the size of the surgeries. Next, the evaluation function is applied, which aims to go through the constraints and then add up the total wastes of each shift. Each waste value is obtained by subtracting the standard time from the surgeries, considering a turnover time of 17 minutes (except for the last surgery), at the time of the assigned shift, and at the end, the total waste is returned to the solution obtained. For each one of the iterations a complete set of waste values is obtained to be compared among alternative solutions.

The iterations end in the limit assigned in the optimisation method, which will return the best value of the solutions obtained. In the end, the satisfactory solution is chosen from the two optimisation methods tested.

4.4 Evaluation

The evaluation phase intends to promote a benchmarking among the obtained models. SMOreg and REPTree models, generated with the top 15 attributes, presented the lowest relative absolute error, i.e. 6.22% and 7.82% of the occurrences, respectively. The relative root quadratic error percentages are 7.59% and 9.26%, respectively.

Initially, it becomes evident that a large amount of data made it difficult to reach the business objective, since the models obtained were complex and large, despite high accuracy values. In this way, two iterations were performed in which only the TOP 10 and TOP 15 of the most relevant attributes for the analysis of the Orthopedics speciality were tested, with satisfactory results. In these two methods, only the characteristics that were documented for the assignment of TEMPO5 (room time) were checked, and no significant changes were obtained. Finally, it was possible to get a model with the lowest percentage of error, enabling the prediction of the duration of each surgery.

In this phase, results of the optimisation methods used to solve the problem were also evaluated. The steps performed in the optimisation block were satisfactory, since there was a significant reduction of the shifts, allowing the placement of all the surgeries without the occurrence of problems, from the preparation of the data until the modelling, being able to minimise the waste of the shift effectively and efficiently. The platform is validated by comparing hill climbing and simulated annealing methods, always choosing the best solution between the two solutions found, namely the solution that presents the least total waste of the shifts for each speciality.

Results

5.1 Analysis of the results obtained in the Prediction Block

Prediction Block includes the trained models as described above. Analysing the error percentages of each model pemits to conclude that the algorithm of Support Vector Regression (SMOreg) is more efficient (because it is the one that presents the lowest percentage about 6% relative absolute error). Table 4 presents an example containing ten surgeries of the orthopaedic speciality, the expected duration time, predicted duration and error.

Table 4 - Prediction sample of the duration of orthopaedic surgeries (time in seconds)

Surgery	Actual	Predicted	Error
1	8100	8049.027	-50.973
2	7020	7127.73	107.73
3	15300	15304.194	4.194
4	5100	5165.485	65.485
5	9000	9026.354	26.354
6	4200	4267.64	67.64
7	9300	9331.48	31.48
8	10560	10761.003	201.003
9	6960	7042.839	82.839
10	7860	7721.515	-138.485

5.2 Analysis of the Results obtained in the Optimisation Block

For the dataset of surgeries and shifts, only the results of two specialities, namely the one with the smallest and the one with the highest number of surgeries, will be discussed, because it is enough to understand the minimisation of the wastes. The speciality with higher duration times of surgery is vascular surgery, which contains 796 surgeries (3456938 seconds) to be distributed for a total of 19389000 seconds, corresponding to 569 possible shifts. Two types of optimisation have been applied to solve this problem: global optimisation for one year; and distributed optimisation for each month of a year. In this method, the optimisation was performed for each of the 12 months that have pendent surgeries, using as a dimension the capacity of surgeries of a respective speciality for the respective month. For orthopaedic a satisfactory result has been attained using of the initial solution generated sequentially, yielding a total loss smaller. The results for each month are as in Table 5.

Table 5 - Monthly sequential optimisation of waste for vascular surgery

	Total waste
January	127089.205125
February	134182.871451
March	171413.244737
April	95961.169928
June	5401.727598
July	23751.340097
August	9194.226731
September	9194.226731
October	29114.226731
Total	605302.200000

For this speciality, it can be concluded that the goal was achieved through monthly optimisation. It was possible to distribute all surgeries by the available shifts, even if they were of a large number and having reduced their shifts by more than half, which means that is possible to introduce more surgeries even though within these restrictions. After analysing the results, the method of optimisation process would be the most profitable for the scheduling of surgeries in the orthopaedic speciality. Nevertheless, for a smaller number of surgeries the optimisation method that offers a better result is the annual optimisation method, precisely because of the ability to make greater management of all possible combinations along a year (which could never happen with large datasets). Despite of these results, it is strongly recommended to use monthly optimisation method, precisely because it is the one that would best fit any dataset.

Conclusions

The use of an ABI approach combining predictive capacities with modern optimisation methods reduces the unforeseen delays in the events of a given organization or institution. To increase the effectiveness of the service, an ABI platform was created using the R environment to facilitate the surgery scheduling task. A dataset of a hospital containing data on surgeries and shifts was used to test the platform. The results are promising and make this approach an efficient and effective ABI event scheduling platform, adaptable to any organisation or institution that needs to schedule large lists of events. It can result in a reduction in unforeseen delays and an increase in the effectiveness of that service, by minimising the waste of shifts. However, and considering the temporal limits and the lack of attributes of the dataset it was not able to address some relevant points that were left for future work, namely: the prioritisation of the the events; the implementation of a graphical interface;the Implementation of a system of analysis of the results through tables, to facilitate the visualisation of the distribution of the events. The approach should be tested considering more complex and abundant data.

Acknowledgments

This work has been supported by FCT – Fundação para a Ciência e Tecnologia within the Project Scope: UID/CEC/00319/2019.

References

1. Michalewicz, Z., Schmidt, M., Michalewicz, M., & Chiriac, C. (2007). Adaptive business intelligence: Three case studies. Studies in Computational Intelligence, 51, 179–196.
2. Michalewicz, Z., Schmidt, M., Michalewicz, M., & Chiriac, C. (2006). Adaptive business intelligence. Adaptive Business Intelligence. https://doi.org/10.1007/978-3-540-32929-9
3. Pregibon, D. (1996) Data mining. Statistical Computing and Graphics Newsletter, 7, p. 8.
4. Gorunescu, F. (2011). Data mining: Concepts, models and techniques. Intelligent Systems Reference Library, 12. https://doi.org/10.1007/978-3-642-19721-5
5. Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning. Elements, 1, 337–387. https://doi.org/10.1007/b94608
6. Jiawei, H., Kamber, M., Han, J., Kamber, M., & Pei, J. (2012). Data Mining: Concepts and Techniques. San Francisco, CA, itd: Morgan Kaufmann.
7. Witten, I. H., Frank, E., & Hall, M. A. (2005). Data Mining Practical Machine Learning Tools and Techniques. Data Mining. https://doi.org/0120884070, 9780120884070
8. Cortez, P. (2014). Modern Optimisation with R. Guimarães: Springer.
9. Azevedo, A. and Santos, M. F. (2008); KDD, SEMMA and CRISP-DM: a parallel overview. In Proceedings of the IADIS European Conference on Data Mining 2008, pp 182–185.
10. Peixoto, R., Ribeiro, L., Portela, F., Filipe Santos, M., & Rua, F. (2017). Predicting Resurgery in Intensive Care - A data Mining Approach. In Procedia Computer Science (Vol. 113, pp. 577–584). https://doi.org/10.1016/j.procs.2017.08.291
11. Coelho, D., Miranda, J., Portela, F., Machado, J., Santos, M. F., & Abelha, A. (2016). Towards of a Business Intelligence Platform to Portuguese Misericórdias. In Procedia Computer Science (Vol. 100, pp. 762–767). https://doi.org/10.1016/j.procs.2016.09.222