Data management for an anaesthesiology department optimization

A. Rabasa[1], A. F. Compañ[2], A. J. Agulló[3], J. J. Rodríguez-Sala[1],
L. Santamaría[1] & L. Noguera[1]

[1]*Operations Research Center, Universidad Miguel Hernández de Elche,
Spain*
[2]*Department of Pathology and Surgery, Universidad Miguel Hernández
de Elche, Spain*
[3]*Anaesthesiology Service, Hospital de San Juan, Alicante, Spain*

Abstract

This paper is part of a Study in Progress which presents the Data Management main guidelines for a live project designed to optimize the anaesthesiology department of a public hospital in Spain. The main goal is the detection of dispensable pre-surgery tests (e.g. chest radiography, blood test or electrocardiogram), if there is a low correlation in the contrast between a patient's clinical history and incidences in the operating theatre. The paper's contribution consists of two parts. Part I shows the development of a data base web application to facilitate data entry of both the anaesthesiology and the operation data itself. The design of the web data base forms and the expert (anaesthetist) specifications are shown. Part II shows how to choose and design the classification rules as the most appropriate Data Mining method for this goal and explains how such a method must be parameterized. Thus, from the historical and clinical data of each new patient, the system could advise about the suitability (or not) for certain expensive and invasive pre-surgery tests (as chest radiography can be). By applying this software, the Hospital will be able to improve the scheduling of critical resources that are often very expensive and cause delays in operation waiting lists. Finally, as future research line, this research aims to extend the proposed method to other medical departments in order to achieve lower costs, avoid invasive tests and waiting lists, in conditions where it is statistically possible.

Keywords: Data Mining, forecasting, optimization, medicine, anaesthesiology, classification rules.

1   Introduction

Nowadays there is a heated debate in the medical community in Spain (and part of Europe) on whether to carry out or not all the preoperative tests that are currently performed (chest radiography, blood test or electrocardiogram). Some of them are expensive and invasive (chest radiography), generate often unnecessary waiting lists and can be expensive for the regional government. Furthermore, the absence of a standard form to collect information leads to very difficult preoperative management of data and it is almost impossible to draw conclusions about the relationships between medical history, evidence and incidences in the operating theatre.

The decentralization and lack of standardization of data (among different hospitals) and even the non-digital format of historical data, greatly complicates their systematic analysis. Therefore, in order to obtain a comprehensive analysis designed to show which tests are dispensable (and under what conditions), firstly it is necessary to design and implement a flexible tool, specifically designed according to the requirements and specifications from the professionals. This tool should collect data that can later be analysed using appropriate techniques.

This paper presents the current status of a research project supported by the Bancaja-UMH grant: "Minería de Datos sobre Análisis Preoperatorios para Servicios Hospitalarios Quirúrgicos y de Anestesia" (Data Mining with Presurgery Tests for Surgery and Anaesthesiology Hospital Departments). The project, as a whole, has two main objectives:

—   In the short term: To provide a web database to standardize data entry
—   In the medium term: To reduce waiting lists, avoid invasive tests and save costs for public health, by analysing data to extract rules of conduct in the database.

This paper focuses on the development of the data base web application to facilitate data entry of both the patient's pre-surgery data and the operation data itself and the selection and design of the classification rules as the most appropriate Data Mining method for surgery incidences forecasting.

In section 2, Methodology, the application specification and design guidelines are presented (Part I), and also a formal reasoning of the classification rules as the most appropriate method for the further analysis (Part II). In section 3, Project implementation, both the functional procedure within the Anaesthesiology department and advances on computational complexity of the author's algorithms are described. Finally, section 4 presents some practical conclusions.

2   Methodology

2.1 Part I: input data specification and forms design

The problem specification requires the inclusion and organisation of 205 attributes related to different groups: patients' personal, medical antecedents, complementary explorations, suggested techniques and instrumental procedures,

and others. There are 42 categorical, 20 numerical, 4 dates, 106 enumerated and 33 multi-valued attributes. All of them are distributed into 21 web forms (on the left side frame in Figure 1): Personal data (Datos personales), Surgery (Intervención quirúrgica), Surgery and anesthesiology antecedents (Antecedentes quirúrgicos y anestésicos), Medical antecedents (Antecedentes médicos), Treatment (Tratamiento actual), Physical examination (Exploración física), Assessment of airway (Evaluación de la vía aérea), Face mask (Mascarilla facial), Complementary examination (Exploraciones complementarias), Interconsultation (Interconsulta), Risk evaluation (Evaluación del riesgo), Anaesthesia (Anestesia), Instructions (Instrucciones), Observations and signature (Observaciones y firma), Intraoperative (Interoperatorio), Anesthetic drugs and adjuvants (Fármacos anestésicos y coadyuvantes), Monitoring (Monitorización), Management of airway (Manejo de la vía aérea), Local anaesthesia (Anestesia regional), Postoperative (Postoperatorio) and Incidences (Incidencias).

The forms were designed by considering the final user specifications, according to the most demanding usability constraints. Patient's data are integrated from different hospital data bases (see Figure 1).

Figure 1:   Personal data form.

Besides, as shown in Figure 2, some visual facilities have been introduced into the interface design. So, the anaesthesiologist must only click on the button radio associated to the graph.

Figure 2:   Assessment of airway form.

Finally, after all attributes have been completed (at their appropriate forms), the operating incidences must be entered on the Incidences form (Figure 3).

Figure 3:   Incidences form.

A total of 205 attributes are grouped by the themes suggested by the experts over 21 forms that allow the most appropriate treatment for each type of data.

Table 1 shows how the variables with different types are distributed into 21 forms.

Table 1:   Attribute types distribution into forms.

Form (attrib)	Text	Numb	date	Enum. (2)	Enum. (3-6)	Enum. (2-n)	Multi-valued
F-0 (2)	2						
F-1 (12)	3	7	1	1			
F-2 (9)		1	2	2	1	1	2
F-3 (3)							3
F-4 (8)						4	4
F-5 (1)							1
F-6 (9)	7			1	1		
F-7 (12)		2		5	5		
F-8 (9)				9			
F-9 (21)	15			5			1
F-10 (7)	1			6			
F-11 (14)				9	2	1	2
F-12 (4)						4	
F-13 (5)	1			1			3
F-14 (4)	1			1		2	
F-15 (15)	2	4	1			4	4
F-16 (4)							4
F-17 (14)				13		1	
F-18 (13)	4			4	3	2	
F-19 (28)	5	6		8	4	2	3
F-20 (5)	1			2	1		1
F-21 (6)				1			5
Total	**42**	**20**	**4**	**68**	**17**	**21**	**33**

2.2 Part II: Selecting and designing the most suitable forecasting method

There are a wide variety of data mining applications for forecasting in Medicine. For example, to predict digestive disorders (Gorzalczany and Gradzki [1]); diabetic disorders (Mugambia *et al.* [2]); and to achieve accurate diagnoses in ophthalmology (Shi *et al.* [3]); in cardiovascular diseases (Gamberger *et al.* [4]) and (Podgorelec *et al.* [5]); or even to manage and plan haemodialysis treatments (Kusiak *et al.* [6]). Of course, the main medical objective: detection and prevention of cancer, has also been approached by using Data Mining techniques: for prostate cancer (Tahir and Bouridane, [7]); cervical cancer (Ho *et al.* [8]); and breast cancer (Polat *et al.* [9]) and (Kohli *et al.* [10]), for example.

On the other hand, the range of possible forecasting techniques used also varies widely. Neural Networks are often used in combination with other techniques, mainly Decision Trees (Nilsson and Sollenborn [11]). They are also combined with Genetic Algorithms (Gorzalczany and Gradzki [1]) where optimal sample simulations become necessary. It is worth mentioning the wide range of papers that show Decision Trees (themselves and their corresponding Rule Sets) as the most suitable predictive method for medical problems. The algorithm C4.5 has been successfully applied on numerous occasions: (Chan *et al.* [12]); (Tahir and Bouridane [7]) and (Polat *et al.* [9]) for numerical target variables; or ID3 for categorical antecedents and class variables.

In this case the experts want to find significant rules that reveal combinations of the previous patient status and tests with possible incidences in the operating theatre. So we face a forecasting problem where the unique target variable (incidences in the operating room) is a multi-valuated one. Thus we will need a classification tree algorithm in order for such a rule set to be built.

Only 20 attributes are numerical, and they are easy to be discretized according to the guidelines set by the anaesthesiologists and the surgeons. Therefore, an ID3 family algorithm will be used.

The RBS algorithm (Reduction Based on Significance), presented by the authors in Abadía *et al.* [13], orders and classifies the rules of a Classification Rule Set into significance regions, according to their corresponding antecedent support and rule confidence values.

Antecedent rule support, *sup(A)*, is defined as follows:

$$sup(A) = \frac{[[A]]}{\left| RS_{A \to C} \right|} = \frac{[[A]]}{n} \tag{1}$$

where [[A]] is the number of possible antecedents in a Rule Set and $\left| RS_{A \to C} \right|$ is the Rule Set cardinality.

Rule confidence, $conf(A \to C)$, is defined as follows:

$$conf(A \to C) = \frac{[[A \to C]]}{[[A]]} \tag{2}$$

where $[[A \to C]]$ is the number of rules with antecedent *A* and consequent *C*.

The RBS algorithm has been tested in different scenarios over very large data sets (more than 2 Gb), providing successful results through ordered and reduced classification rule sets. However, in this paper we face a very "broad" problem because 204 attributes (with several possible values) must be taken into account as antecedents and an automatic attribute selection procedure becomes absolutely necessary. Therefore, the cited RBS algorithm has been updated (RBS+) to include a feature selection function to filter the most correlated attributes with the incidence class variable (Figure 4).
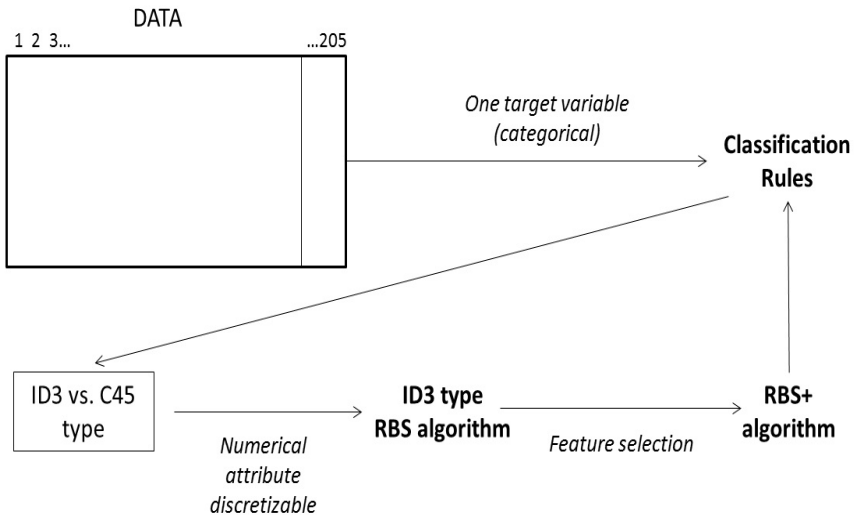

Figure 4:   Selecting the most suitable forecasting method.

3   Current project implementation

The web application has been recently installed in the hospital computers and linked to the hospital data bases and life data pick up is starting. Around 10 records (surgery information) per day are expected to be collected. While real data is increasing, it will be used, as a seed, to simulate a bigger input data set. A control point on the project will analyze such mixed data in September, 2013. Mid 2014, a significant sample of life data is expected to be collected and then the anesthesiologists and surgeons will be provided with the first classification rule set.

Meanwhile, improvements are being developed in the algorithm RBS+ regarding the RAM efficient management and computational cost reduction. Also, Genetic Algorithm techniques are being developed to improve the automatic feature selection procedure and new quality measures of the input sample are being tested in order to achieve the best accuracy as possible.

4 Conclusions and future research lines

This paper leads to three main conclusions that are as follows:

1) Some pre-surgery tests could be avoided, but it is absolutely necessary to confirm it statistically, by finding very low correlations between tests and incidences on the operating room.

2) Due to the nature of the variables and the unique target variable (incidences), in this case, the most suitable forecasting method is the classification trees that provide classification rule sets that can be reduced and ordered by the RBS algorithm.

3) Because of the amount of variables to be taken into account, a very friendly interface design becomes necessary.

The authors plan three fronts on which to continue their research. One refers to the improvement of the interface, the second one refers to the generalization of the tool and the last one involves improvements to computational models for Data Mining. To be exact, the future lines are as follows:

1) The development of a graphical output module is planned for the rule systems visualization on a friendly interface which is easy and practical for the experts.

2) As a part of the project, a user's management module will be provided to enable a generalized use of the application. Thus, it will be applied to other hospital departments, allowing differentiation between different user roles.

3) New pre-pruning heuristics will be developed to achieve optimal rule sets generation.

References

[1] Gorzalczany, M.B. and Gradzki, P. Computational Intelligence In Medical Decision Support - A Comparison Of Two Neuro-Fuzzy Systems. *Proc. ISIE'99*, 1999.

[2] Mugambia, E.M., Hunterb, A., Oatleyd, G. and Kennedy, L. Polynomial-Fuzzy Decision Tree structures For Classifying Medical Data. Elsevier. *Knowledge Based Systems*, 17, pp. 81-87, 2004.

[3] Shi, W., Wahba, G., Wright, S., Lee, K., Klein, R. and Klein, B. LASSO-Pattern Search Algorithm With Application To Ophthalmology, 2006.

[4] Gamberger, D., Lavrac, N. and Krstacic, G. Confirmation Rule Induction And Its Applications to Coronary Heart Disease Diagnosis and Risk Group Discovery. IOS Press. *Journal of Intelligent and Fuzzy Systems* 12, 1, pp. 35-48, 2002.

[5]   Podgorelec, V., Kokol, P., Stiglic, M.M., Hericko, M. and Rozman, I. Knowledge Discovery With Classification Rules In A Cardiovascular Dataset. Elsevier. *Computer Methods and Programs in Biomedicine* 80 Suppl. 1, pp. 39-49, 2005.

[6]   Kusiak, A., Dixon, B. and Shah, S. Predicting Survival Time For Kidney Dialysis Patients: A Data Mining Approach. Elsevier. *Computers in Biology and Medicine*, 2004.

[7]   Tahir, M.A. and Bouridane, A. Novel Round-Robin Tabu Search Algorithm For Prostate Cancer Classification And Diagnosis Using Multispectral Imagery. IEEE-Inst. Electrical Electronics Eng. *IEEE Transactions on Information Technology in Biomedicine* 10, 4, pp. 782-793, 2006.

[8]   Ho, S.H., Jee, S.H., Lee, J.E. and Park, J.S. On Risk Factors for Cervical Cancer Using Induction Technique. Elsevier. *Expert Systems with Applications* pp. 27, 97–105, 2004.

[9]   Polat, K., Sahan, S., Kodaz, H. and Gunes, S. A New Classification Method For Breast Cancer Diagnosis: Feature Selection Artificial Immune Recognition System, FS-AIRS. Springer-Verlag. *Advances in Natural Computation 2, Proc. Lecture Notes in Computer Science* 3611, pp. 830-838, 2005.

[10] Kohli, R., Krishnamurti, R. and Jedidi, K. Subset-Conjunctive Rules for Breast Cancer Diagnosis. Elsevier. *Discrete Applied Mathematics* 154, pp. 1100 – 1112, 2006.

[11] Nilsson, M. and Sollenborn, M. Advancements and Trends in Medical Case-Based Reasoning: An Overview of Systems and System Development. American Association for Artificial Intelligence. Malardalen University. Technical Report, 2004.

[12] Chan, A.L., Chen, J.X. and Wang, H.Y. Application of Data Mining to Predict the Dosage of Vancomycin as an Outcome Variable in a Teaching Hospital Population. Dustri-Verlag. *International Journal of Clinical Pharmacology and Therapeutics* 44, 11, pp. 533-538, 2006.

[13] Abadía, R., Almiñana, M., Escudero, L.F., Pérez-Martín, A., Rabasa, A. and Santamaría, L. A Rule Reduction Algorithm Based on Significance Measure. WIT Press. *Transactions on Information and Communication Technologies*. 40, pp. 63-72, 2008.