Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 98 (2016) 528 – 533

Procedia Computer Science

The 3rd International Symposium on Emerging Information, Communication and Networks (EICN 2016)

Resurgery clusters in Intensive Medicine

Ricardo Peixoto[a], Filipe Portela*[a,b], Filipe Pinto[d,e], Manuel Filipe Santos[a], José Machado[a], António Abelha[a] and Fernando Rua[c]

[a]*Algoritmi Research Centre, University of Minho, Portugal;* [b]*ESEIG, Porto Polytechnic, Portugal*
[c]*Intensive Care Unit, Centro Hospitalar do Porto, Portugal ;* [d] *Universidad de las Fuerzas Armadas, ESPE, Quito, Ecuador*
[e]*Instituto Politécnico de Leiria, Leiria, Portugal*

Abstract

The field of critical care medicine is confronted every day with cases of surgical interventions. When Data Mining is properly applied in this field, it is possible through predictive models to identify if a patient, should or should not have surgery again upon the same problem. The goal of this work is to apply clustering techniques in collected data in order to categorize re-interventions in intensive care. By knowing the common characteristics of the re-intervention patients it will be possible to help the physician to predict a future resurgery. For this study various attributes were used related to the patient's health problems like heart problems or organ failure. For this study it was also considered important aspects such as age and what type of surgery the patient was submitted. Classes were created with the patients' age and the number of days after the first surgery. Another class was created where the type of surgery that the patient was operated upon was identified. This study comprised Davies Bouldin values between -0.977 and -0.416. The used variables, in addition to being provided by Hospital de Santo António in Porto, they are provided from the electronic medical record.

Peer-review under responsibility of the Program Chairs

Keywords: Clustering, Data Mining, Intensive Care Units, INTCare, Re-intervention, Intervention;

1. Introduction

This study is inserted in the INTCare project. This study intends to use clustering techniques to characterize the resurgery patients. This is a pioneering approach because nothing has been done previously related to this work in intensive care. This work is different from all the other previously performed, since it addresses cases where patients needed further surgery (again to the same problem). This study aims to improve the performance of Intensive Care

* Corresponding author. Tel.: +0-000-000-0000 ; fax: +0-000-000-0000 .
E-mail address: cfp@dsi.uminho.pt

Peer-review under responsibility of the Program Chairs
doi:10.1016/j.procs.2016.09.072

Units (ICUs) assisting health professionals in their decision-making regarding their patients. The dataset used in this project was provided by ICU of Centro Hospitalar do Porto. A resurgery is performed when a patient needs to be operated upon again to a given problem. This has only recently been analyzed as a problem, which means that there is not an extensive work in this area. The goal of this work is to identify health problems and characteristics of resurgery patients in order to prevent these cases from happening again. This study presents some interesting results with the best value of Davies-Bouldin being -0.416. The negative value provides the fact that the tool used to assess this value always deliver results with negative value. This work is divided in five sections. The first one, Introduction, where the basic ideas for this work are presented, the second is the Background, where the problem will be defined and the theory behind the work, the third is the description of the study, d the methods and tools used are described, and understanding the data. In the fourth chapter, named Discussion, some views on the results of this study are presented. Finally, in the last chapter are presented some conclusions and and the future work.

2. Background

2.1. Intensive Care and INTCare

In the 90s, experts realized that the available knowledge was not enough to solve complex real-life problems [1]. The change was then induced, where the main focus was the collection of knowledge directly from data, using intelligent data analysis. After 90 years, the approach has gained more interest, especially in the medical field due to the large volume and complexity of clinical data [2]. Intensive Medicine is categorized as a multidisciplinary field of medical science that specifically addresses the prevention, diagnosis and treatment of potentially reversible acute disease conditions in patients presenting imminent bankruptcy or established in one or more vital functions [3]. First, they have to make a diagnosis and design their treatment plans to improve the patients conditions [4]. However, doctors in intensive care units have to make decisions that are even more challenging, such as life support treatments. In this field appeared INTCare. The INTCare main goal is the development of an intelligent system able to predict clinical events [2] and can be accessed anytime and anywhere. Currently, INTCare is a pervasive intelligent decision support system based in a fully automated knowledge discovery process and in. It also uses the agent paradigm[5, 6]. The values automatically collected are validated by the nurses in order to ensure their quality [7].

2.2. Interventions and re-interventions

In the past, surgical interventions were seen as the last resort for certain diseases. However, due to the evolution of science, surgery also started to be used as a treatment. With the development of new techniques and the evolution of knowledge, surgical methods have been given preference, driven by advances in anesthesia, antisepsis, radiology and blood transfusion and the use of bone transplants or prostheses [8]. Surgical procedures are part of their everyday life in health according to WHO, in a survey conducted in 56 countries In 2004, the percentage of surgeries was 20% [9]. Another study [10] showed that only in industrialized countries the rate of complications of surgeries was 3% to 6%, the mortality rate between 0.4% and 0.8%. Three-quarters of surgical procedures are performed in only one third of the world population[9]. According Cunha[11], surgical intervention can be divided into three phases: preoperative, intraoperative, and postoperative. In this case the incidence of trauma and complications are high and where there is a huge backlog of untreated surgical diseases[12]. In a study provided by Lafortune et al [13] it is possible to observe that the most frequent intervention in Portugal is to cataracts. A resurgery occurs when a particular patient in this section need re-intervention to a given problem. Regarding resurgery made, in intensive medicine, is not possible to obtain any statistics since this is an area which is still under investigation.

It should be stated that this problem was recently studied, since this is an aspect that only recently begun to be given as a problem. In practical terms, there is no DM practical work related to surgical re-interventions, however it was already developed data mining work in the medical field, for example work related with surgical interventions, however, related surgical re-intervention is a pioneering study as it addresses cases where patients have been intervened and requires a new intervention. Some work in this area were already developed in ICUs [14,15,16]. Other works also were developed in data mining area, however not being included in the area of intensive care [17].

3. Study description

3.1. Methods and Tools

Clustering is a common descriptive task, which seeks to identify a finite set of categories or groups to describe the data [18]. According Cheeseman and Slutz [19], the categories can be mutually exclusive and exhaustive or consist of a richer representation, as hierarchical or overlapping categories. Clustering can be seen as research into natural groupings of objects based on similarity measures, i.e., similar profiles [1]. This is a process of partitioning a heterogeneous population homogeneous in various sub-groups [20]. In Intensive medicine there is few works using clustering techniques. One of the examples is the categorization of readmitted patients [21]. The CRISP-DM (Cross Industry Standard Process for Data Mining) was followed as the data mining methodology. The life cycle of a data mining project [22] is divided into six phases: Business understanding; Data understanding; Data preparation; Modeling; Evaluation and Deployment. The tools used were Oracle SQL Developer, for data exploration and preparation, Orange and RapidMiner for building scenarios and clustering data. In Orange the following tests were used: Silhouette, Inter-cluster distance and Distance to centroids. In Rapid Miner, k-means algorithm and Davies-Bouldin Index were used. According to Petrovic [23], the Davies-Bouldin index of the clustering is combined with the centroid diameters comparison between clusters. In the computation of the Davies-Bouldin index, the centroid linkage is used as the inter-cluster distance [23]. During the development of the project two different tools, which enabled at the same time to explore two different evaluation metrics were used. The used scenarios in both tools are the same, which means that the end result quite reliable, since they are available two different methods for comparison.

3.2. Business Understanding, Data understanding and Data preparation

The idea of this work is providing information for professionals in this area in order to understand which patients that usually needs a resurgery. Data mining techniques were used with the goal to characterize these patients.

Table 1 and Table 2 present the distinct values, the percentage of cases the average minimum and maximum values, in case of numeric variables, of the dataset used in this work. Notice that this study used real data acquired from CHP databases. Some data changes and corrections to eliminate the fields with errors were performed in order to improve the data quality. The used strategies have gone through the data standardizing, calculating the age of the patients according to their date of birth, as well as creating classes with these ages and the different types of surgery, that is, all surgeries belong to a specialty, and neurology, cardiology, or other. An important data changes made was to replace all the false and true for 0 and 1 respectively. In the table 1 it is possible see that the variable "sex" varies between 1 and 2, being the male and female respectively. This variable was received with these values.

Table 1 – Two Distinct Values Variables

Variable	Distinct Values	Value 0	Value 1
Sex	2 (1 or 2)	35.35%	64.65%
CardiovascularDisease	2 (0 or 1)	96.74%	3.26%
RespiratoryDisease	2 (0 or 1)	95.41%	4.59%
NeurologicalDisease	2 (0 or 1)	99.62%	0.38%
OncologicDisease	2 (0 or 1)	96.26%	3.74%
Diabetes	2 (0 or 1)	99.53%	0.47%
Transplant	2 (0 or 1)	93.78%	6.22%
Immunosuppressed	2 (0 or 1)	96.49%	3.51%
Hypocoagulability	2 (0 or 1)	99.53%	0.47%
OtherDisease 1	2 (0 or 1)	93.49%	6.51%
Sepsis	2 (0 or 1)	90.61%	9.39%
CardiovascularInstability	2 (0 or 1)	93.49%	6.51%
RespiratoryInstability	2 (0 or 1)	97.61%	2.39%
HypovolemicShock	2 (0 or 1)	93.87%	6.13%
OtherDisease2	2 (0 or 1)	97.99%	2.01%
Urgency	2 (0 or 1)	41.57%	58.43%

Table 2 - Other variables

Variable	Distinct value	Average	Minimum	Maximum
DurationOfSurgery	51	-	01:48 (1)	06:15 (60)
NDaysAfterFirstSurgery	39	25.109	0	179
Age	-	64.391	25	95
SurgeryCategory	10	-	Least (Pneumonia (6))	Most (Global Surgery (512))
ClassOfAges	14	-	Least (25-30 (4))	Most (54-55 (265))

3.3. Modeling

This phase was focused in achieving the business goals through the use of clustering techniques. In order to achieve the expected results nine scenarios that allow characterize the re-intervened patients were developed. The first set of attributes created was Case Mix, which contained all the variables from the table used. The next scenarios were created by using some criteria like the attributes targeting patients that had surgery, or patients that had neoplasms. Other scenarios (like 4 and 8) were created to understand if some variable had positive or negative impact in the final results. With these criteria 9 scenarios (S1-S9) were created. At Rapid Miner (DMM1), 9 models were generated and at Orange (DMM2), 27 models were generated. These models can be represented by:

DMM1 = {9 Scenarios, 1 Technique, 1 Sampling Method, 1 Representation Method, 1 Target}

DMM2 = {9 Scenarios, 3 Techniques, 1 Sampling Method, 1 Representation Method, 1 Target}

S1:{MainDiagnosis, SurgeryType, CardiovascularDisease, RespiratoryDisease, NeurologicalDisease, DurationOfSurgery, OncologicDisease, Diabetes, Transplant, Immunosuppressed, Hypocoagulability, WoundInfectionorBurn, OtherDisease1, SurgeryCategory, Sepsis, ClassOfAges, CardiovascularInstability, ClassDaysAfterFirtsSurgery, RespiratoryInstability, HypovolemicShock, OtherDisease2, Sex, Urgency}

S2:{MainDiagnosis, SurgeryType, CardiovascularDisease, RespiratoryDisease, NeurologicalDisease, DurationOfSurgery, OncologicDisease, Diabetes, Transplant, Immunosuppressed, Hypocoagulability, WoundInfectionorBurn, SurgeryCategory, Sepsis, ClassOfAges, CardiovascularInstability, ClassDaysAfterFirtsSurgery, RespiratoryInstability, HypovolemicShock, Urgency;

S3:{CardiovascularDisease, RespiratoryDisease, NeurologicalDisease, DurationOfSurgery, OncologicDisease, Diabetes, Transplant, Immunosuppressed, Hypocoagulability, SurgeryCategory, Sepsis, ClassOfAges, CardiovascularInstability, ClassDaysAfterFirtsSurgery, RespiratoryInstability, HypovolemicShock, Urgency}

S4:{CardiovascularDisease, RespiratoryDisease, NeurologicalDisease, OncologicDisease, Diabetes, Transplant, Immunosuppressed, Hypocoagulability, Sepsis, ClassOfAges, CardiovascularInstability, ClassDaysAfterFirtsSurgery, RespiratoryInstability, HypovolemicShock, Urgency, NDaysAfterFirstSurgery}

S5:{Diabetes, SurgeryType , CardiovascularInstability, HypovolemicShock, RespiratoryDisease, Hypocoagulability, Urgency, Immunosuppressed, Transplant, Sepsis, OncologicDisease, NeurologicalDisease, RespiratoryDisease, CardiovascularDis}

S6:{ClassOfAges, ClassDaysAfterFirtsSurgery, SurgeryCategory}

S7:{ClassOfAges, SurgeryCategory, NDaysAfterFirstSurgery}

S8:{ClassOfAges, SurgeryCategory, CardiovascularDisease, CardiovascularInstability, RespiratoryInstabilit, Urgency, HypovolemicShock, Sepsis, Hypocoagulability, Immunosuppressed, Transplant, Diabetes, OncologicDisease, NeurologicalDisease, RespiratoryDisease, NDaysAfterFirstSurgery}

S9:{CardiovascularDisease, CardiovascularInstability, RespiratoryInstability, Urgency, HypovolemicShock, Sepsis, Diabetes. Hypocoagulability, Immunosuppressed, Transplant, OncologicDisease, NeurologicalDisease, RespiratoryDisease}

In the table 3 are presented the settings used to carrying out the tests for scenarios.

Table 3 - Algorithm settings

Algorithm	Setting	Value
k-means	K	2 to 8
	Max Runs	10
	Max optimizations steps	300
	Measures type	Numeric Measures
	Numerical Measures	Silhouette; Inter-cluster distance and Distance to centroids

To find the most correct number of clusters with k-means technique was used Davies-Bouldin Index. In the table 4 are shown all realized scenarios, where the best results can be observed.

Table 4 - Orange and Rapid Miner results

Scenario	Silhouette	Inter-cluster distance	Distance to centroids	Davies-Bouldin Index	Number of clusters
5	0.90	1.990	0.2300	-0.977	10
7	0.81	61.990	6.2117	-0.416	3
8	0.81	61.029	6.0280	-0.417	3
9	0.89	1.430	0.2100	-0.977	10

3.4. Evaluation

The scenario were observed in detail in order to understand which models presents the best results in respect of defined metrics. The chosen scenario can be observed in greater detail in the table 4. This phase was focused first on the assessment of the results provided by the use of k-means. At this stage it was possible to observe which cluster presents the best results and what the best scenario to resolve the problem of re-interventions is. To build the table 4 the models with best results were used. It is important to understand that the choice of the best cluster was based on the value of silhouette given by the used application. Analyzing in more detail the table 4 it is possible to observe that the majority of the re-interventions cases are distributed by the clusters 3 and 10. In figure 1 it is possible to observe the separation of Urgency variables with values 0(False) and 1(True) between the clusters 1 and 2. As it is possible to see, it is an important variable to categorize re-intervened patients to ICU. This information is grouped by the patients with an urgency admission. In the figure 2 it is possible to observe the distribution of the patients with Respiratory Disease. As can be seen in the figure 2, patients with Respiratory Disease should go to cluster 2 while patients who do not have this health problem should go to the cluster 1.

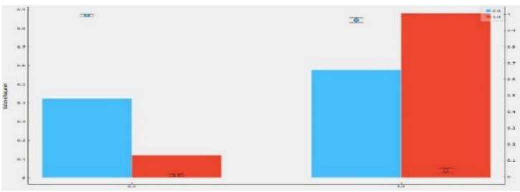

Figure 1 - Distribution of Urgency for each cluster

Figure 2 - Distribution of Respiratory Disease for each cluster

4. Discussion

After making the evaluation of the models it was possible observe that the number of days after the first surgery is one of the attributes with most impact. This variable can change the patient’s situation, i.e., if he is or not re-intervened.

Attributes such as number the days after the surgery have great relevance in Resurgery predictive of patients. Analyzing figure 3 it is possible observe that there is a good separation between the three clusters (C1, C2, C3). In the case of patients with value of days after the surgery between 0 and 20 days they are almost in C1.

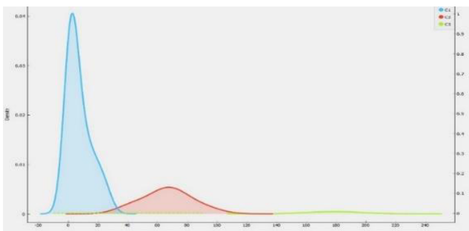

Figure 3 - Distribution of DaysAfterFirstSurgery for each cluster

5. Conclusion and Future Work

When analyzing the related work it was possible to observe that there is not current practical work that can fit in this project. All the work done in prior scientific terms proved to be quite useful, since it has been used to realize not only as the Intensive Care Unit act as the remaining points of the section number two. In the ICU, the number of patients and the number of variables available for each patient is huge. This situation means that the number of possibilities is quite high, however this study has provided a useful aid to characterize the re-intervened patients.

The developed clusters cannot guarantee that patients will re-intervened but they provide information on which the clinical situation and the types of patients that should be considered by physicians. With this work new variables were created to predict resurgery patients. The number of days that the patient is re-intervened after the first surgery is one of the most important attributes. As mentioned in previous sections, to measure the quality of the created clusters and to define what best represent re-intervened patients, its necessary to have in mind the silhouette values, inter-cluster distance and distance to centroids. Combining all the obtainable information from the set standards and the knowledge and experience of physicians and nurses, it should be possible to these health professionals to identify which patients are usually re-intervened.

Acknowledgments

Work supported by Fundação para a Ciência e Tecnologia within the Project Scope UID/CEC/00319/2013.

References

1. Portela, F., Santos, M., Vilas-Boas, M., Rua, F., Silva, Á.M.. Real-Time Intelligent Decision Support in Intensive Medicine. KMIS 2010 - Knowledge Management and Information Sharing,. pp 44-50. ISBN: 978-989-8425-30-0. Valência, Spain. SciTePress. (2010).
2. Filipe Portela, Manuel Filipe Santos, Álvaro Silva, José Machado, António Abelha and Fernando Rua. Pervasive and Intelligent Decision Support in Intensive Medicine - The Complete Picture . LNCS) 2014, pp 87-102. ISBN: 978-3-319-10264-1. Springer. (2014)
3. Silva ÁJBM da. Modelos de Inteligência Artificial na análise da monitorização de eventosclínicos adversos, Disfunção/Falência de órgãos e prognóstico do doente critico. 2007.
4. Hanson CW, Marshall BE. Artificial intelligence applications in the intensive care unit. 2001;29(2):427–435.
5. Cardoso L, Marins F, Portela F, Santos M, Abelha A, Machado J. The next generation of interoperability agents in healthcare. *Int J Environ Res Public Health*. 2014;11(5):5349–5371. doi:10.3390/ijerph110505349.
6. Portela F, Vilas-Boas M, Santos MF. Improvements in Data Quality for Decision Support in Intensive Care. In: Szomszor M, Kostkova P, eds. *Electronic Healthcare: Third International Conference, eHealth 2010, Casablanca, Morocco, December 13-15, 2010, Revised Selected Papers*. Berlin, Heidelberg: Springer Berlin Heidelberg; 2012:86–94. doi:10.1007/978-3-642-23635-8_11.
7. Santos, Manuel · Portela, Filipe. Enabling Ubiquitous Data Mining in Intensive Care - Features Selection and Data Pre-Processing. ICEIS 2011 - International Conference on Enterprise Information Systems. pp 261-266. ISBN: 978-989-8425-53-9. China. SciTePress. (2011).
8. Maziero ECS. Avaliação da implantação do programa de cirurgia segura num hospital de ensino. 2012.
9. Weiser TG, Regenbogen SE, Thompson KD, et al. An estimation of the global volume of surgery: a modelling strategy based on available data. *Lancet*. 2008;372(9633):139–44. doi:10.1016/S0140-6736(08)60878-8.
10. Gawande AA, Thomas EJ, Zinner MJ, Brennan TA. The incidence and nature of surgical adverse events in Colorado and Utah in 1992. *Surgery*. 2016;126(1):66–75. doi:10.1067/msy.1999.98664.
11. Cunha V. Laserterapia e ultrassom no tratamento pós-operatório da cirurgia plástica de abdominoplastia : Revisão de literatura. 2012:1–14.
12. Debas HT, Gosselin R, Mccord C. Chapter 67 Surgery. 2004.
13. Lafortune G, Balestat G. Comparing activities and performance of the hospital sector in Europe: how many surgical procedures performed as inpatient and day cases? *Organ Econ Co-operation Dev Dir Employment, Labour Soc Aff*. 2012;(December):33–36. Available at: http://www.oecd.org/health/Comparing-activities-and-performance-of-the-hospital-sector-in-Europe_Inpatient-and-day-cases-surgical-procedures.pdf.
14. Pereira JJR. Modelos de Data Mining para multi-previsão : aplicação à medicina intensiva. 2005.
15. João M. C. Gonçalves, F. P., Manuel F. Santos, Álvaro Silva, José Machado, António Abelha. Predict Sepsis Level in Intensive Medicine - Data Mining Approach. Advances in Intelligent Systems and Computing Volume 206, 2013, pp 201-211. Springer. (2013).
16. Braga P, Portela F, Santos MF, Rua F. Data mining models to predict patient's readmission in intensive care units. *6th Int Conf Agents Artif Intell ICAART 2014*. 2014;1(Im):604–610. Available
17. Rui Veloso, Filipe Portela, Manuel Filipe Santos, Álvaro Silva, Fernando Rua, António Abelha, José Machado . Categorize readmitted patients in Intensive Medicine by means of Clustering Data Mining. (IJEHMC) . IGI Global. (2016). (accepted for publication).
18. Jain AK, Dubes RC. Algorithms for clustering. *Cap IV-Cluster Validity*. 1988:143–222.
19. João M. C. Gonçalves, Filipe Portela, Manuel F. Santos, Álvaro Silva, José Machado, António Abelha, Fernando Rua. Real-time Predictive Analytics for Sepsis Level and Therapeutic Plans in Intensive Care Medicine. IJHISI - Volume 9, Issue 3, pp 36-54. IGI Global. (2014)
20. Filipe Portela, Rui Veloso, Sérgio Oliveira, Manuel Filipe Santos, António Abelha, José Machado, Álvaro Silva and Fernando Rua. Predict hourly patient discharge probability in Intensive Care Units using Data Mining. Indian Journal of Science and Technology. Volume 8, Issue 32. ISSN: 0974-6846. Indian Society for Educat. (2015)
21. C. Alberto, D.S. Ferreira, H. Daniel, N. Fernandes, V. Lúcia, O Data Mining na Compreensão do Fenómeno da Dor : Uma Proposta de Aplicação, (2006).
22. Wirth R, Hipp J. CRISP-DM: Towards a Standard Process Model for Data Mining. 2000.
23. Petrovic S. A Comparison Between the Silhouette Index and the Davies-Bouldin Index in Labelling IDS Clusters. *Tthe 11th Nord Work Secur IT-systems, Nord 2006*. 2006:53–64. Available at: https://xp-dev.com/svn/b_frydrych.../silhuetteIndexRegulaStopu.pdf.