CENTERIS - International Conference on ENTERprise Information Systems / ProjMAN - International Conference on Project MANagement / HCist - International Conference on Health and Social Care Information Systems and Technologies, CENTERIS / ProjMAN / HCist 2017, 8-10 November 2017, Barcelona, Spain

Knowledge Discovery from Surgical Waiting lists

Cristiana Neto[a], Hugo Peixoto[b], Vasco Abelha[b], António Abelha[b] and José Machado[b]*

[a]University of Minho, Campus Gualtar, Braga 4710, Portugal
[b]Algoritmi Research Center, University of Minho, Campus Gualtar, Braga 4710, Portugal

Abstract

Methods for knowledge discovery in data bases (KDD) have been studied for more than a decade. New methods are required owing to the size and complexity of data collections in administration, business and science. They include procedures for data query and extraction, for data cleaning, data analysis, and methods of knowledge representation. The part of KDD dealing with the analysis of the data has been termed data mining. Data mining is the process of pattern discovery and extraction where huge amount of data is involved. Data mining has been used intensively and extensively by many organizations. In healthcare, data mining is becoming increasingly popular, if not increasingly essential. Data mining applications can greatly benefit all parties involved in the healthcare industry. In this work is presented an approach for the use of data mining in the context of waiting lists for surgery, namely for predicting the type of surgery (programmed or additional) for a record in the list.

Keywords: Knowledge Discovery in Databases, Data mining, Surgery waiting list, Decision Support Systems, Classification

* Corresponding author. Tel.: +351 253 604 430; fax: +351 253 604 471.
E-mail address: jmac@di.uminho.pt

1. Introduction

Among the huge and varied number of fields that currently exist, data are being collected and stored at a tremendous rate[1]. More and more organizations place the needs that information provides on other entities but often these organizations, such as healthcare institutions, find themselves with a huge amount of data but without truly valuable information[2].

So, in order to be able to assist humans in extracting useful information (knowledge) from the growing amount of digital data, there is an urgent need for a new generation of computational theories and new tools. These theories and tools are the subject of the emerging field of knowledge discovery in databases (KDD)[3]. The KDD is the term used refer a multistep process of discovering useful knowledge from data and one particular step in this process is the application of data mining algorithms for extracting patterns (models) from data. Recently, data mining has invoked significant interest in its applications to healthcare.

The healthcare environment is usually information rich but knowledge poor. However, data mining techniques can be applied to create a knowledge rich healthcare environment. The specific part of healthcare that this work focuses on is the analysis of information about waiting lists for surgery in a hospital.

Waiting lists have a strong influence on access to health care and may mean that there is an inability to satisfy basic human health needs. The waiting lists increased in Portugal in the last decades and one of the main complaints of patients is the high waiting times for surgery, which can often be years. It's important to emphasize that waiting lists for surgery tend to be more pronounced in countries that combine health insurance[4].

In 2004, an information system was implemented (SIGIC – Integrated System for the Management of Enrolled Patients for Surgery) within Portuguese health care institutions and led to a better planning and programming of the institution's activity, reducing waiting times through management.

One of the measures taken with the implementation of the SIGIC in order to promote the recovery of surgical waiting lists was the creation of additional production. So nowadays, two types of surgery can be distinguished: programmed/base and additional. The type of surgery that a patient is receiving is very important in terms of hospital management, especially with regard to its financial management.

Due to this importance, it would be interesting to predict the type of surgery of a record in the waiting list given certain attributes. This prediction can be done using data mining algorithms (classification or regression). In our case, as the objective is to predict the type of surgery, the most adequate algorithms are the classification ones. For this process, several tools can be used and the chosen was the Weka tool. This paper presents a contextualization on the subject of knowledge extraction, particularly the data mining technique and the results obtained, as well as its discussion.

Regarding the structure of this paper, it is divided in six major sections. After the current introduction, the second section presents related work and the background concepts of the given problem which is waiting lists. The methodologies, data and methods used are then described in next section. Section four describes the steps of the Cross Industry Standard Process for Data Mining (CRISP-DM) methodology that was followed during the DM process. In section five the obtained results are discussed. Finally, section six includes conclusions and future work.

2. Background and Related Work

As defined by Fayyad, Piatetsky-Shapiro, and Smyth[5], "*Knowledge discovery in databases is the non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data.*" A KDD is a multistep process that automatically examines data to determine the types of relationships, using a variety of techniques from statistical, artificial-intelligence, or machine-learning algorithms[5]. The quality of health services can be improved using clinical systems with a built-in decision support system, where KDD is very relevant, reducing error and costs.

Portela et.al.[6] presented a way how INTCare, an Intelligent Decision Support Systems (IDSS) developed in the intensive care unit of the *Centro Hospitalar do Porto*, could accommodate new functionalities. This paper presented an approach to the KDD procedure in order to enable a pervasive, online and real-time processing of data in ICU,