Chapter 13

Healthcare Delivery in a Hospital Emergency Department

Joseph Twagilimana
University of Louisville, USA

ABSTRACT

The outcome of interest in this study is the length of stay (LOS) at a Hospital Emergency Department (ED). The Length of stay depends on several independent clinical factors such as treatments, patient demographic characteristics, hospital, as well as physicians and nurses. The present study attempts to identify these variables by analyzing clinical data provided by electronic medical records (EMR) from an emergency department. Three analysis methodologies were identified as appropriate for this task. First, data mining techniques were applied, and then generalized linear models and Time series followed. In spite of the fact that Data Mining and Statistics share the same objective, which is to extract useful information from data, they perform independently of each other. In this case, we show how the two methodologies can be integrated with potential benefits. We applied decision trees to select important variables and used these variables as input in the other models.

BACKGROUND

Medical data form a collection of observations about a patient. Among records on each patient are the patient identification (name and/or medical record), the parameters being observed, the values of the parameters, and the time of observation. Some parameters such as complaints, diagnoses, charges or treatments are in textual format. Some others such as age, blood pressure, and temperature are in numerical format. There are also pictures such as x-rays and CT scans. All of these characteristics make medical data analysis a challenging one. Krzysztof J. Cios and G. William Moore (2002) report the major points of the uniqueness of medical data:

- Heterogeneity or volume and complexity contained in the data. They can include Physician’s interpretation (unstructured free-text English), sensitivity and specificity

DOI: 10.4018/978-1-61520-723-7.ch013
Healthcare Delivery in a Hospital Emergency Department

analysis, poor mathematical characterization, and lack of canonical form (imprecision, lack of standard vocabulary).

- Ethical, legal, and social issues: Because of a fear of lawsuits directed against Physicians, unnecessary lab tests are ordered. This makes the analysis of the data difficult because there is no association or pattern with the patient’s condition.

- Statistical Philosophy: Assumptions underlying Statistical methods such as normality and homogeneity of variance are constantly violated in medical data. Medical data are primarily patient care oriented rather than a research source, unless they are from designed experiments. The size of the patient sample is often too small for data mining techniques, which require a considerable amount of data for training/validation methodology. Medical data are highly noisy and often contain missing fields. Some data mining techniques such as neural networks do not handle missing data. There are also issues with clustering in large dimensional medical data. A major concern is how to incorporate medical domain knowledge into the mechanism of clustering. Without focus and at least partial human supervision, one may end up with results that do not make sense.

- Special status of medicine: Medicine has a special status in life. The outcomes of medical care can be life-or-death. Medicine is a necessity, not merely an optional luxury, pleasure, or convenience. Any medical data analysis must take into consideration that there is no room for errors. Any results must be accompanied by a certain confidence level to help decision makers make the right choices.

The ideas presented in the above paragraph were borrowed from the article written by Krzysztof J. Cios and G. William Moore (2002). We recommend this article to beginners in medical data analysis.

SETTING THE STAGE

The data to be analyzed were collected using the Ibex Pulse Check, a comprehensive and well-integrated emergency department information system (EDIS). Ibex Pulse Check is a fast-charting system that gives both emergency nurses and physicians the ability to view the medical record as it is being created (Pulse Chek, n.d.) The data were collected for a 3-month period in a hospital emergency department. It had originally 12 variables and 3345 observations. As any medical data, the datasets contain variables in text format (complaints, diagnosis, charges), numerical variables (age) and some other identifier such as patient name, Physician’s initial, and Nurse’s initial. Variables were recoded to maintain the privacy of patients, Physicians and Nurses. New variables were created as they were needed for the preprocessing and analysis phases of the study. A counter variable, visits, that counts the patients as they enter the emergency room, and a variable count, which accumulates the number of patients per hour were created. The variable Time, which specifies the hour interval time in which the patient was seen by the triage nurse, was created from the variable triage.

The variable, charges, which specifies the items for which the patient was charged, was used to create clusters of patients. It is believed that patients with similar symptoms and diagnoses are treated in a similar manner. Those clusters were created using the SAS Text Miner. Text mining can be used to group patients with similar complaints, diagnoses and charges in the same cluster. The new variable showing the clusters is labeled “CLUSTER”. The following table gives a summary description of the variables. The data have been de-identified according the HIPAA recommendations. Table 1 gives a brief description of the data.