An International Journal
JSCDSS
http://www.jscdss.com
Vol.5 No.4 August 2018: 26-32
Article history:
Accepted 12 August 2018
Published online 13 August 2018

Journal of Soft Computing and Decision Support Systems

E-ISSN: 2289-8603

Prediction of Patients' Mortality during Hospitalizations

Mojtaba Zare [a,*], Janusz Wojtusiak [a], Mehrbakhsh Nilashi [b]

[a] Department of Health Administration and Policy, College of Health and Human Services, George Mason University, Fairfax, VA 22030-4444, USA
[b] Faculty of Computing, Universiti Teknologi Malaysia, 81310, Skudai, Johor, Malaysia

*** Corresponding authors email addresses**: mzare@gmu.edu

Abstract

In this study, we predict patients' mortality in the session that a patient is hospitalized. We focus on both lab results and vital signs collected in the first 24 hours of a patient admission to predict his/her mortality within that hospitalization. We use MIMIC-III dataset for data analysis and building predictive models. We include only patients with age of 18 or above resulting in a sample of 38,578 patients. Independent variables include patients' demographic information, lab results and vital signs. The dependent variable is whether the patient dies within that hospitalization. We use Weka 3.8 for data analysis and model building. After randomly splitting the data into 80 and 20 percent, then we use the 80 percent of the data for feature selection as well as training the prediction models. As a result, we construct four prediction models using Bayes Network, Logistic Regression, Naïve Bayes and Random Forest. After constructing these models, we test them on the remaining 20 percent of the data. We use Receiver Operating Characteristic (ROC) area, precreation and recall values to compare the accuracy of these models with each other. Result showed that the highest ROC area belongs to the Bayes Network at 0.824 followed closely by Logistic Regression at 0.817 for the test set.

Keywords: Mortality, MIMIC-III, Classification algorithm, Prediction models

1. Introduction

Mortality prediction of hospitalized patients is a crucial task. It is important to accurately predict the mortality of a patient because it benefits both the patient and health care resources (Awad et al., 2017; Lee et al., 2015). Over the past recent years, machine learning models have been proposed to predict mortality of patients with different type of diseases such as breast and lung cancer (Martín-Sánchez et al., 2016; Malvezzi et al., 2015), and chronic kidney disease (Haapio et al., 2017; Vijayarani et al., 2015). However, most of previous studies have focused on building data mining models to predict patient morality after patients' ICU admission or after discharge (Awad et al., 2017; Wojtusiak et al., 2017). Hence, early mortality prediction regardless of patients' admission to ICU or patients discharge remains an open challenge. In this study, we predict patients' mortality in the session that a patient is hospitalized. We focus on both lab results and vital signs collected in the first 24 hours of a patient admission to predict his/her mortality within that hospitalization.

2. Method

2.1 Dataset

In this study, we used MIMIC-III (Medical Information Mart for Intensive Care III) dataset for data analysis and building prediction models. MIMIC-III is a large freely database containing information of 46,520 patients who stayed in critical care units of the Beth Israel Deaconess Medical Center between 2001 and 2012. The database contains patients' information such as "demographics, vital sign measurements made at the bedside, laboratory test results, procedures, medications, caregiver notes, imaging reports, and mortality (both in and out of hospital)" (Johnson et al., 2016).

2.2 Variables

In this study, we included only patients with age of 18 or above. This sample consists of 38,578 patients. Because patients may have more than one admission, we randomly chose one admission for each patient. After conducting an extensive examination of data variables, we included patients' demographic information, lab results and vital signs as independent variables. Patient demographic information consisted of age, race and gender. We focused on both lab results and vital signs variables collected in the first 24 hours of a patient admission to predict his/her mortality within that hospitalization. Lab result included information about ALBUMIN, BILIRUBIN, CREATININE, CHLORIDE, GLUCOSE, HEMATOCRIT, HEMOGLOBIN, PLATELET, POTASSIUM, SODIUM, BUN, WBC and WBC_count rate of the patients. Vital