

ORIE 4741 Project Proposal: Inpatient Charges Prediction

Group Members: Siyao Gu, Yiling Jiang, Yingying Zheng

September 21, 2017

1 Motivation

A recent study found that medical bills are the major cause of more than 60 percent of all personal bankruptcies in the U.S.A. Despite the Obamacare, the skyrocketing medical treatment expenses and ever growing health insurance premiums put many stresses on the general public, especially on the middle class which makes up about 53 percent of the whole population in the U.S.A. Hence our team is dedicated to study the healthcare system with a goal to help society have a better understanding of the cost associated with different medical treatments. This project is aimed at predicting inpatient charges based on the data from New York State hospitals in 2012, and we hope to detect potential medical bill frauds and abuses for healthcare insurance companies.

2 Question

What are the factors in determining the hospital charges for inpatients at discharge? Are the charges for inpatient treatment predictable?

3 Data and Approach

3.1 Data Description

The data that we are going to examine are obtained from New York State data portal:

<https://health.data.ny.gov/Health/Hospital-Inpatient-Discharges-SPARCS-De-Identified/u4ud-w55t>

The dataset contains 2012 New York State hospital discharge level details on facility name, hospital county, patient characteristics, diagnoses, treatments, services, and charges. We decide to predict the inpatient charge based on features such as hospital area, patient characteristics, length of stay, treatment, severity of illness, and so on. In the meantime, we also plan to study the relationships between the features and the charge and develop insights on how heavily each feature makes an impact on the charge.

3.2 Assumptions

Our first assumption is that the number of frauds and abuses within the dataset is significantly small so that it is acceptable to ignore them and treat all the inpatient charges as expenses without frauds and abuses. We believe that most of the charges made are reasonable. Therefore, the existence of frauds and abuses in the data will not affect our prediction model, and we can trust our model to make fair predictions.

We also assume that inpatient charges are not subject to changes in market and healthcare policies throughout the years. Under this assumption, we believe our prediction model is still feasible for future years.

3.3 Selection of Data

After examining the dataset, we found there are more than one hundred types of treatments each of which has a unique ID. Due to the difficulty of including every single treatment in the model, we plan to start the model with a subset of data containing the most frequent treatments.