Modeling Healthcare Quality via Compact Representations of Electronic Health Records

Ankia Dande,Pragati Shinde ,Anjali Soundale,Ashwini Borod ,Prof Sonali Potdar

Abstract

Nowadays health care has attracted many researchers' attention. Heath care domain has huge amount of impact in data warehouse. With increasing number of patients, data is increasing.it became very difficult to manage detail of every patient manually.to solve that problem, it is needed to maintain that record electronically. This Electronic Health record has proven to be very effective in case to improve the quality of health care system. With adapting automated system in the medical area, everything is becoming more reliable. This paper couple EHRs with the advanced machine learning tools to predict three important parameters of healthcare quality. More specifically, this paper has described how to learn low-dimensional vector representations of patient conditions and clinical procedures in an unsupervised manner, and generate feature vectors of hospitalized patients useful for predicting their length of stay, total incurred charges, and mortality rates.

This paper has proposed Naïve Bayes classifier. We have used Dataset of 1000 records, which contains a discharge details of the patients those have done different kinds of congestive heart surgeries.When user selects Disease, City and Hospital, on the basis of given dataset, Naïve Bayes classifier finds the probability and predicts the Length of Stay, total incurred charges, and mortality rates of the patient. Proposed system helps users to choose the hospital that will fit in their needs and their financial constraints.

Keywords: Electronic Health Records; healthcare quality; embedding models; neural language models.

Introduction

Nowadays due to changed living standards, living habits, the incidence of heart disease is increasing. so many people don't have aware about which hospital offers which kind of services .Even they are worried about the hospital charges ,cause of that they get confused while selecting the hospital for major surgeries .

However, computing this value upfront is not a trivial task, as pricing of health care services vary significantly among different Hospitals even for the most common procedures. Each hospital takes into account many parameters before charging a patient.

For these reasons, many economists, employers and health plans are advocating for providing the price quote of health care services as a way to encourage consumers to choose low-cost, high-quality providers and to promote competition based on the value of care.

Machine learning has proved to be very effective in the past decades. In the area of health care, machine learning has improved and fastens the diagnosis process. With machine learning, the estimation of each process can be predicted accurately. This system is made for the hospital to improve the quality of their treatment. With this system, the patient can easily get the prediction of its duration of staying and the price patient need to spend in drugs and treatment. This paper also predicts the survival rate of the patients in that hospital that were suffering from the same disease. These measures can be used to help hospitals identify potential problem areas that might need further studies and provide the opportunity to assess quality of care inside hospitals using administrative data found in typical discharge records

Proposed System provides information about health care services in advance. This paper has proposed Naïve Bayes classifier. We have used Dataset of 1000 records, which contains a discharge details of the patients those have done different kinds of congestive heart surgeries.When user selects Disease, City and Hospital, on the basis of given dataset, Naïve Bayes classifier finds the probability and predicts the Length of Stay, total incurred charges, and mortality rates of the patient. Proposed system helps users to choose the hospital that will fit in their needs and their financial constraints.

Goal of the System

Motivated by above problems we planned to design the application that provides information about health care services in advance.

Main goal of proposed system is to predict Length of Stay, total incurred charges, and mortality rates of the patient before operation. It will help patients to select the hospital, which provides the best service in affordable rate.

SYSTEM ARCHITECTURE

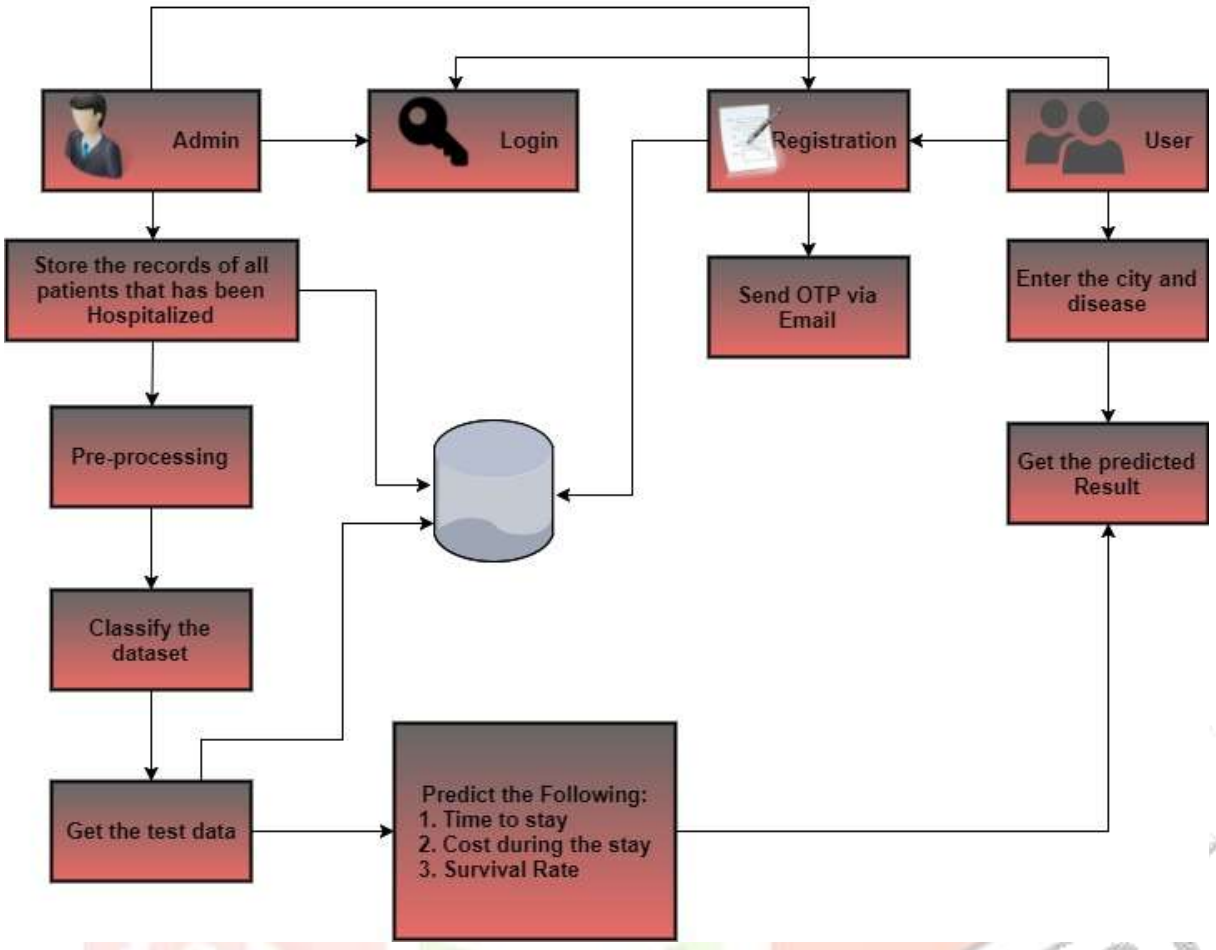

Fig: System Architecture

As shown in above Architecture, it contains following Modules

1. Register and Login

With the help of this module user will register and login into the system.

2. Admin Login

With the help of this module admin will login into the system **.**Admin has authority to add and view the hospitals into the system.

3. Add Hospital

Admin uses this module to add hospital into the system**.**

4. View Hospital

This module shows hospitals those already added into the system.

5. Predication

After login user will select city, disease and hospital on the basis of that system will predict LOS, Survival Rate and Total hospital charges.

6. Ambulance

This module shows list of ambulance number as per the hospital.

7. Review

With the help of this module user can give review about the hospitals. Reviews given by the users will help other users to select the hospital, which provides the best service in affordable rate.

8. Logout

This module helps user and admin to logout from the system.

Related Works

There is need to mine clinical data in order to obtain actionable insights to improve healthcare services. That concept is called as data-drivenhealthcare. In this approach practitioners havebeen addressing various problems aimed to improve healthcarequality. the main object of this approach is to build a framework to model different aspects of healthcare system.it will help institutions and patients to get significant information about healthcare services .

Some particularly important and impactful applications are aimed towards predictive modelingof health outcomes in terms of diseases, procedures, mortality,and other measures that may have a huge impact on quality ofpatient treatment. The models are used to improve detection ofhigh-risk groups of patients, or detect important effects not takeninto consideration in prior medical treatments.However, the modeling process is very challenging, as healthcareobservational data are often sparse, heterogeneous, and/orincomplete due to different hospital and insurance policies, furtheraggravated by non-standardized physician practices . Theexisting data mining tools are not fully capable of addressing theimportant task of

healthcare modeling, and, in order to makeuse of multifaceted, noisy healthcare data sources, developmentof novel efficient and effective machine learning approaches isrequired.

In this study we address this important problem, and proposed Naïve Bayes classifier. Naïve Bayes classifier finds the probability and predicts the Length of Stay, total incurred charges, and mortality rates of the patient. Proposed system helps users to choose the hospital that will fit in their needs and their financial constraints.

HYPOTHESES

1- We are developing a system that helps a user to predict the Length of stay, Total Cost and Survival rate before operation the patient.

2- User has to select disease, city and hospital ,on the basis of that system will predict the Length of stay , survival rate and total hospital charges .

3- Naïve Bayes algorithm is used to predict theLength of stay , survival rate and total hospital charges. Naïve Bayes algorithm provides increased accuracy in the predication.

4- Naïve Bayes algorithm uses training dataset to find the and predict theLength of stay ,survival rate and total hospital charges.

DESIGN OF THE STUDY

Implementation phase focus over system design objectives. Software implementation is the process of designing, writing, testing, debugging / troubleshooting, and maintaining the source code of computer programs.

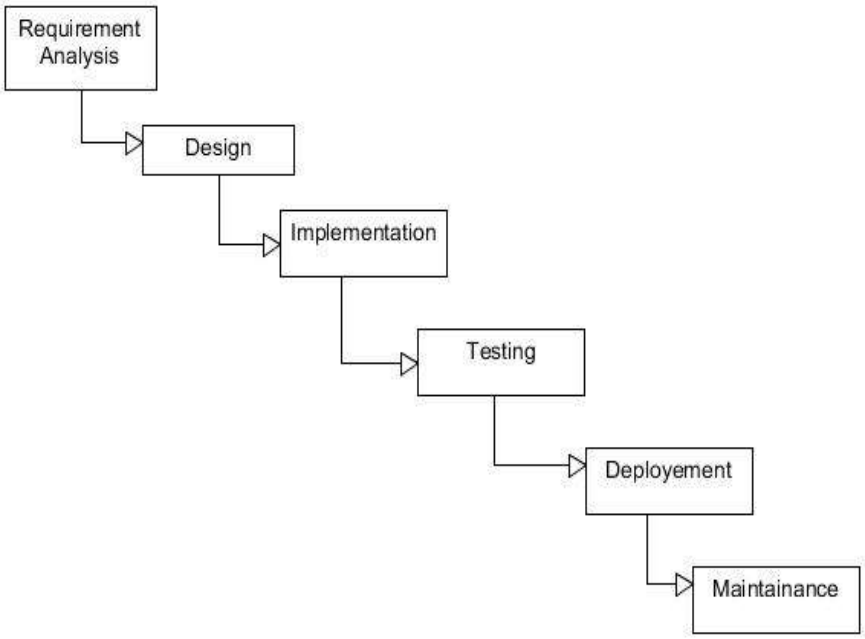

Fig : Software Development Process

With the help of Register and Login module user will register and login into the system.With the help of Adminmodule admin will login into the system .Admin has authority to add and view the hospitals into the system.Add Hospital module is used to add hospital into the system**.** View Hospitalmodule shows hospitals those already added into the system. After login user will select city, disease and hospital on the basis of that system will predict LOS, Survival Rate and Total hospital charges.Ambulancemodule shows list of ambulance number as per the hospital.With the help of Reviewmodule user can give review about the hospitals. Reviews given by the users will help other users to select the hospital, which provides the best service in affordable rate.Logoutmodule helps user and admin to logout from the system.

SAMPLE OF THE STUDY

We are developing a system that provides information about health care services in advance. This paper has proposed Naïve Bayes classifier. We have used Dataset of 1000 records, which contains a discharge details of the patients those have done different kinds of congestive heart surgeries.When user selects Disease, City and Hospital, on the basis of given dataset, Naïve Bayes classifier finds the probability and predicts the Length of Stay, total incurred charges, and mortality rates of the patient. Proposed system helps users to choose the hospital that will fit in their needs and their financial constraints.

TOOLS UDED

Software Requirement:

- Operating System : windows 10
- Application Server : Tomcat 8.0
- Language : Java
- Front End : HTML, JSP
- Database : MySQL
- IDE : Eclipse

Hardware Requirement: The hardware design of the system includes designing the hardware units and the interface between those units.

- Processor - Pentium –IV
- RAM - 1 GB (min)

Hard Disk - 20 GB

STATISTICAL TECHNIQUE USED

We have developed Login and Registration module which manages the user profiles. User gives Disease, city and Hospital name as an input. Along with that a patient's discharge dataset is loaded into the system. This dataset contains personal details of the patients as well as discharge details those have done heart surgery .on the basis of given input system predicts Length of stay, Cost, Survival rate of the patient.

ALGORITHM

Naïve Bayesian Classifier:

Naïve Bayesian classifier works on probability.

Let $\vec{X}$ = {X1,X2…..Xn} be the records of patient

And A1,A2 …..An be their attributes

The classifier needs to predict $\vec{X}$ belongs to the class with the highest a probability, i.e., $\vec{X}$ is predicted to lie in the class Ci if and only if there exists i, such that:

$$P(Ci|\vec{X}) > P(Cj \mid \vec{X}), \text{ for all } 1 \leq j \leq f, \ j \neq i$$

By Bayes's theorem,

$$P(Ci|\vec{X}) = P(\vec{X} \mid Ci)\ P(Ci)/\ P(\vec{X})$$

Where,

$$P(\vec{X} \mid Ci) \approx \prod 1\_(k=1)\wedge n \ [\![P]\!] \ (Xk \mid Ci)$$

1. The first step of this system is to represent the EHR into the low dimensional vector space. To learn the representation for disease and procedure, we will use one objective function as:

$$\mathcal{L} = \sum_{r \in R} \sum_{hi \in r} \sum_{-b \le m \le b, m \ne 0} \log P(h_{i+m} | h_i)$$

Where $\mathcal{L}$ iterates over the whole dataset. Probability $P(h_{i+m}|h_i)$ of observing some "neighbouring" disease/procedure h_{i+m} given the current disease/procedure h_i is defined using the soft-max function as

$$P(h_{i+m}|h_i) = \exp(v_{hi}^T v'_{hi+m}) / \sum_{h=1}^{H} exp(v^T_{hi} v'_h)$$

wherev_h and v'_h are the input and output M-dimensional vector representations of disease/procedure h and hyper-parameter b represents the length of the context for disease records. Note that h can represents either d or p, with H = |D| + |P|.

2. **Patient Visit Representation:**

The above step was to generate a disease-procedure vector. But, this step will only map to the patient's disease and their procedure through their symptoms. Next, we need to create a dataset which will represent the features like LoS, total charges or binary mortality indicator. This can be done by adding both disease and procedure that appeared in that record. Let x_i be patient's feature vector.

$$X_i = \sum_{j=1}^{Di} v_{dij} + \sum_{l=1}^{Pi} v_{pil}$$

3. The problem of predicting first two classes i.e. Length of Stay and total charges can be implemented via linear regression. With simple linear regression we want to model our data as follows:

$$y = B0 + B1 * x$$

This is a line where y is the output variable we want to predict, x is the input variable we know and B0 and B1 are coefficients that we need to estimate that move the line around. The steps of linear regression are as follows:

Step 1: Let x be the input variable and y be the output variable. First, we need to calculate the mean value of x and y. The mean is calculated as:

OUR APPROACH

Proposed method achieves the prediction based on low-dimensional embedding models. The problem with applying this classical linear algebra theory is that in many applications such a simple linear subspace model does not apply. Nonetheless, there are a handful of alternative low-dimensional models appropriate for wide varieties of real-world signals. Before discussing applications to specific healthcare related prediction problems, it is intuitive to introduce neural language models as applied to NLP. These methods take advantage of word order, and assume that closer words in the word sequence are statistically more dependent. Typically, a neural language model learns the probability distribution of the next word given a fixed number of preceding words that act as the context.

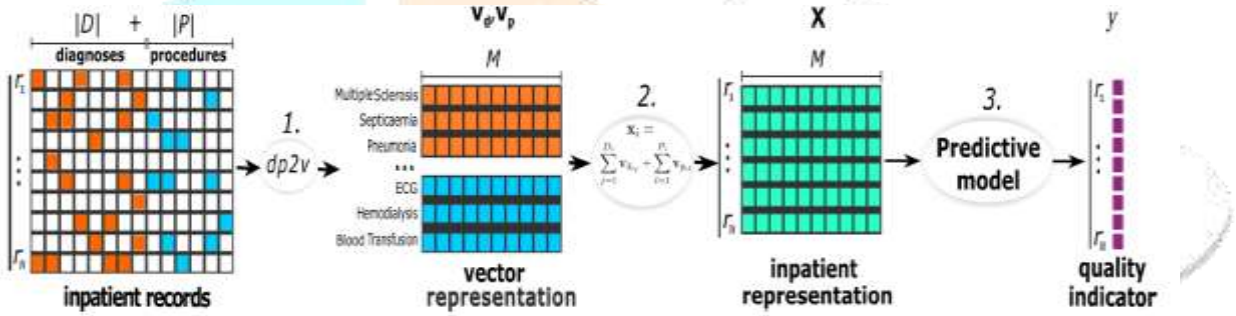

Fig : Pipeline of proposed approach

Experiment Result:

In this section we provide experimental results of three predictive tasks on dataset.

Dataset

We have used Dataset of 1000 records, which contains a discharge details of the patients those have done different kinds of congestive heart surgeries.

We have included following disease categories in the dataset

- Insertion of implantable heart assist system
- Implantation of cardiac resynchronization defibrillator total system (CRT-D)
- Implantation of cardiac resynchronization defibrillator pulse generator (CRT-D)
- Insertion of percutaneous external heart assist device
- Heart transplantation
- Excision destruction or exclusion of left atrial appendage (LAA)
- Aquapheresis
- Automatic implantable cardioverter-defibrillator (AICD)
- Noninvasive programmed electrical stimulation (NIPS)
- Removal of lead(s) [electrode] without replacement
- Endovascular removal of obstruction from head and neck vessel(s)
- Replacement of automatic cardioverter-defibrillator lead(s) only

Prediction of total charges

In this section we address the problem of predicting total charges for a patient per hospital visit. There are more than 100 factors that may influence hospital charges, making the estimation of the exact value a non-trivial problem. We collected the discharge records (total charge) of the patients from different hospital about the patients those have done heart surgery. Whereas an input we take into account diagnosed diseases for a specific patient and a list of procedures and charges that applied. Navies Bayes classifier finds the probability of the charges and predicts the total cost.

Prediction of Length of Stay

The length of stay is one of the most important indicators of quality of a hospital system, and is an important parameter considered when choosing a hospital. And also useful for planning scheduling capacity within a hospital.

Therefore, providing LoS estimation for a specific visit is a very important task. Many hospitals are handling these predictions by reporting the mean length of stay.

In this study we consider a patient that is diagnosed with several diseases, as well as number of days they were there in the hospital, and we account for procedures suggested for this patient in order to estimate the patient's length of stay.

Prediction of inpatient survival

Lastly, we turn our attention to estimating patients mortality, which we use as an ultimate quality indicator of hospital care considered in this study. More specifically, the prediction task was to estimate patient's survival probability, taking into consideration diagnosed conditions and conducted procedures.

We collected the survival information about patients those are diagnosed with several diseases as well diagnosed conditions and conducted procedures. and we trained the system accordingly .

By considering disease, system predicts the survival rate patient

When user selects Disease, City and Hospital, on the basis of given dataset, Naïve Bayes classifier finds the probability and predicts the Length of Stay, total incurred charges, and mortality rates of the patient. Proposed system helps users to choose the hospital that will fit in their needs and their financial constraints.

Future work

The methodology still possesses drawbacks in terms of modelling diseases and procedures embedding. For example, currently the model does not account for the concept of primarydiagnosis and secondary diagnoses, heterogeneity of a disease is not captured well by the given approach and multiple visits of same patients, including readmission, are not included in the modelling process. Modelling longitudinal effects and addressing disease heterogeneity will be the focus of our future work.

Conclusion:

This paper proposes a framework to predict Length of Stay, Survival Rate, and total charges before operating the patient. It uses Naïve Bayes Algorithm to find the probability and predict the result.

This paper has described how to learn low-dimensional vector representations of patient conditions and clinical procedures in an unsupervised manner, and generate feature vectors of hospitalized patients useful for predicting their length of stay, total incurred charges, and mortality rates.

Reference:

1. Patients and policy makers increasingly use rates of surgical mortality to assess hospital performance.
   https://pdfs.semanticscholar.org/4d2d/82742ea1af8cf082846a210096bdbb306563.pdf.
2. The contribution of medical costs to household bankruptcy risk by exploiting plausibly exogenous variation in publicly provided health insurance.

http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.502.6146&rep=rep1&type=pdf.

3. The co-occurrence of medical and job problems was a common theme.

http://www.pnhp.org/PDF_files/MedicalBankruptcy.pdf**.**

4. To assess excess length of stay, charges ,and death attributable to medical injuries during hospitalization.

https://pdfs.semanticscholar.org/321d/72b58af16b49897fc566b379f58205440a41.pdf.

5. The Mining electronic health records an additional perspective data.

https://healthcare.utah.edu/huntsmancancerinstitute/research/labs/population-sciences/smith/images/04_Miningelectronichealthrecords.pdf.