International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Impact Factor (2012): 3.358

A Mining Method to Predict Patient's DOSH

Ruchi Rathor[1], Pankaj Agarkar[2]

[1]Department of Computer Engineering, Dr D.Y.Patil School of Engineering, Savitribai Phule Pune University, India

[2]Professors, Department of Computer Engineering, Dr D.Y.Patil School of Engineering, Savitribai Phule Pune University, India

Abstract: ***Management of hospital resources is a major and composite activity that can highly influence the work rate of the hospital's services. Mostly, resource management is needed when patients are hospitalized, as large amount of resources are under exhaustion at that specific time. Hence, predicting the number of days a patient stays at the hospital can help in organizing the hospital resources. In this paper we propose a prediction model that predicts the Duration of Stay at the Hospital (DOSH) by the patient. We used basic clustering and classification method for the prediction. In this methodology, hidden anomalies and deviations can be brought out which cannot be featured by applying basic clustering, that will give an efficient prediction outgrowth.***

Keywords: clustering, classification, prediction

1. Introduction

One of the most important factor over which the productivity of the hospital's service is totally dependent, is the management of the hospital resources. These resource comprises of beddings, medicines, food etc. management of the hospital resources is required so that (a) resources that are misused and wasted, can be prevented; (b) resource can be utilized efficiently; (c) better estimation of the resources can be performed; (d) future resource demands can be planned efficiently; (e) future appointments can be handled; (f) proper medical services can be served to the patients and (f) can result into high occupancy rate[1] [4]. Hence, managing the resources will enhance the productivity of the hospital and quality of life, of the patient [2].

This paper proposes a DOSH prediction model that will predict the number of days a patient will stay at the hospital and based on that how much amount of resources will be required by the patient will be estimated and reserved for that respective patient. The prediction is made three times, beginning with when patient arrives to consult the doctor, second at the admission of the patient and third based on the condition of the patient. If the condition of the patient progresses as predicted then no further predictions are made but if the condition deteriorates or doesn't thrive, the possible second prediction or further third predictions are made.

2. Related Work

By knowing, the need for Duration of Stay of the Patient in the Hospital exorbitant research been conducted and studied. Since 1960s many prediction models are being created that can predict the Duration of stay (LOS) of the patient in the hospital and their result being compared.

D. H. Gustafson [3] proposed and compared five prediction methodologies out of which three resulted with point estimation based on physician's estimation whereas other two couldn't perform well and gave poor precision as the data collected was inadequate and estimation couldn't be performed.

E. K. Kulinskaya and H. D. Gao [2] investigated the factors that can affect the prediction of the LOS of the patient. The data used was a statistical diagnosed data. As statistical LOS data does not give normal distribution and consist of lots of outliers, they proposed a robust statistical methodology on order to handle outliers and a normal distribution can be formed. To analyze effect of factors on LOS prediction two methods are compared that worked over the data .One is Standard Method: General Linear Models (GML) and Robust Method: Truncated Maximum Likelihood (TML). Out of the two TLM proved to be better estimator of the prediction. But the accuracy of the prediction made is not compared with the actual one.

V. Liu, P. Kipnis, M. K. Gould, and G. J. Escobar [5] predicted LOS based on linear regression model and data set from 17 hospitals with total of 205,177 hospitalizations. In addition, they added Laboratory Acute Physiology Score (LAPS) and Co-morbidity point score (COPS) to linear regression model gave improved outcome.

Ali Azari, Vandana P. Janeja , Alex Mohseni [4] to cut down uncertainty of LOS at hospitals, proposed a multi-tiered data mining approach in which four different data training sets were created out of which three were formed by applying different clustering approaches and one by non-clustering that were analyzed by ten different classification algorithms. Each training set is processed by each classification algorithm forming about forty models. These models were compared based on performance measures of classification algorithms. The performance measures used were accuracy, recall, area under curve kappa statistics and precision. For ranking the models Friedman test was conducted, which concluded that Support Vector Machine and Bnet generated the better predictions. This proposal couldn't give appropriate predictions for outliers and weakly performed for clusters of dynamic shapes and densities. Also, due to presence of anomalies tuples prediction was not truly efficient.

Panchami V U [1] proposed a model in which LOS was predicted such that LOS longer than seven days. The dataset used was statistical data from the hospitals. By using

different classification approaches four non-identical models were devised. They used Naïve bayes (NB), Neural-Network (NN), Support Vector Machine (SVM), also Logistic Regression (LR). They compared accuracy, precision and recall of these models. Also statistical magnitude was calculated by t-test. They used DBSCAN clustering method which acted as forerunner for classification method. Also they said clustering followed by classification gives better outcome as noises are removed only prime data are grouped together. The clusters are formed in order to treat outliers and clusters of dynamic shapes and sizes. This is done by DBSCAN algorithm.

There were three training sets were formed; first training set was created without performing any clustering, second by k-means clustering and third by DBSCAN. In addition, an extra training set also created called test set. These test set and training set does not share common data. The three were compared by three performance measures accuracy, precision and recall. The result is analyzed and ranking is done and conclusion is made. [1]

2.1 System Architecture

The system embodies four modules – data processing, clustering, classification and result evaluation/assessment. Whereas, Classification module is sub-divided into trainer and tester.[1]

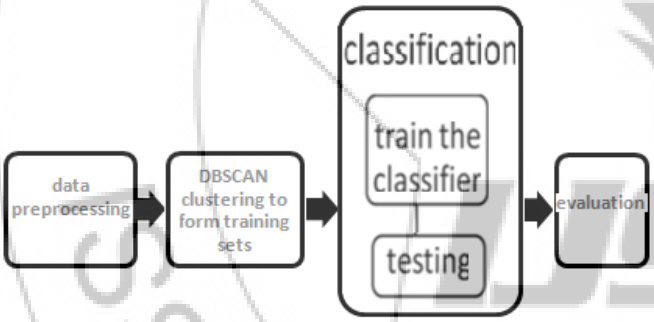

Figure 1: Basic system architecture

Step1: Preprocessing of data
Databases are very much vulnerable to inconsistency, noisiness and absent data. To get the precise outcome these error generating factors should be fixed. To get the precise output the upcoming steps are taken: Data Cleaning, Data Integration and Transformation and Data Reduction.

Step2. DBSCAN
DBSCAN stands for density based spatial clustering for applications with noise. Noises here referred to as outliers contained in the clusters. DBSCAN finds clusters with irregular shapes.

Step3. Classification
Four classifiers are used to train the training sets. They are Naïve Bayes (NB), Neural-Network (NN), Support Vector Machine (SVM), also Logistic Regression (LR).

Step3. Evaluation
The result was evaluated by comparing the performance measures i.e accuracy, recall and precision of each of the models.

Conclusion

From the analysis it is concluded that model formed by DBSCAN clustering and classified by SVM gave the efficient prediction. But, this proposal does not say anything about the inconsistency or the anomalies in the database. [1]

3. Proposed System

This study proposes a new clustering-classification approach for the prediction of Duration of Stay at the Hospital (DOSH) of the patient. Clustering followed by Classification will extract the significant data from the huge database and also remove the noisy data from the data. DBSCAN do removes the noisy and dense datasets with efficiency but it has three main hindrances, as:

(1) DBSCAN burdens the user to provide the parameters that will in finding the next clusters. That is, parameters like MinPts and epsilon-neighborhood is to be initially known. [7].

(2) The calculation of distance metric between the two points (x,y) for the function (a, epsilon) where a is a point and epsilon is the neighborhood; give good clusters. DBSCAN works well for in a Euclidian space. But for multi-dimensional data this distance metric be bad. [6].

(3) When working with Hierarchical datasets DBSCAN will not respond as competently. [6].

We argue that the clusters formed by our Clustering-Classification approach will also consider the hidden anomalies and will give the precise prediction. Before Clustering the anomalous data will be separated so that we get more representative training set.

The proposed system consists of three modules. With a elaborated functionality of data convertor module, data mining module and the evaluator module.

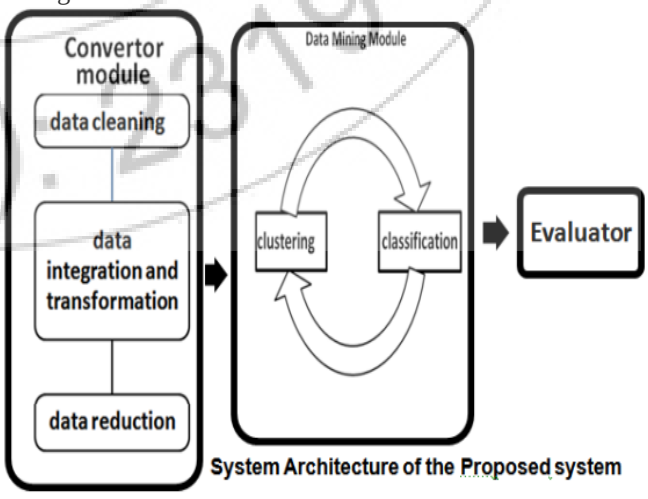

Figure 2: Proposed System Architecture

(1) Convertor Module
Convertor Module is nothing but the data preprocessor. As huge databases are very much prone to noisy, missing and inconsistent data there is a need to preprocess this data. These data are of very low quality and a low quality data will give a low quality in mining results. Thus, there is a need to enhance the quality of the data in such databases. For this purpose, data preprocessing is required.

Preprocessing consists of three steps:
(a) Data Cleaning,
(b) Data Integration and Data Transformation,
(c) and Data Reduction

(2) Data Mining Module
Data mining module has two recursive steps: clustering and classification. Clustering will form the training sets which will be classified further and this classified cluster will again be clustered, to form a refine training set this way Hidden anomalies will be highlighted.

(3) Evaluator Module
The evaluator module will analyze the prediction made by the system and update it as required.

3.1 System Functionality

Patient will enter its symptoms and medical history to the system. The system will predict the DOSH of the patient. Based on these details and prediction report, required resources are reserved for that specific patient. The resources are reserved in advance but not allocated to the patient. This is done, because this specific prediction is not finalized until the domain specialist or doctor predicts their DOSH for the patient. If any specific resource is not available for that time, it can be arranged soon.

This prediction report will be send to the doctor. Doctor will conduct some tests and finally will predict DOSH based on the initial prediction report and the tests conducted. Now based on this prediction the resources are allocated to the patients and admitted.

After admitting, the complete track of patient's condition is kept. This information of the patient will be continuously updated in the hospital database after every certain interval of time. Based on this, DOSH is predicted. Further prediction is totally based on requirement. For example, if the condition does not improve or degrades further a new prediction is made. And further treatments are done on its basis. But if condition improves as expected, no further prediction is required.

Hence, in our system the prediction is done three times.
(1) First, at the time of visit.
(2) Second, during admission based on tests
(3) And third, after admission based on current condition of the patient.

4. Conclusion & Future Work

To predict Duration of Stay at the Hospital (DOSH) of the patient we applied the concepts of Data Mining in our system. Also hidden anomalies were included. The prediction is made three times based on some medical factors.

Future work: our future studies will comprise of security of patient's private sensitive data to prevent exposure to unauthorized people. Also to can have anomaly detection at the starting of convertor module, so that a more precise training sets can be formed.

References

[1] Panchami V U and N.Radhika "A novel approach for predicting the length of Hospital stay with DBSCAN and supervised classification algorithms," in 2014 IEEE.
[2] E. K. Kulinskaya and H. D. Gao, "Length of stay as a performance indicator: robust statistical methodology," IMA JOURNAL OF MANAGEMENT MATHEMATICS, vol. 16, no. 4, pp. 369–381, 2005.
[3] D. H. Gustafson, "Length of stay prediction and explanation, "Health Services Research, vol. 37, no. 3, pp. 631-645, 2002.
[4] Ali Azari, Vandana P. Janeja , Alex Mohseni, "Predicting Hospital Length of Stay (PHLOS) : A Multi-Tiered Data Mining Approach," in 2012 IEEE 12th International Conference on Data Mining Workshops, pp.17-24, 2012.
[5] V. Liu, P. Kipnis, M. K. Gould, and G. J. Escobar, "Length of stay predictions: Improvements through the use of automated laboratory and comorbidity variables," Medical care, vol. 48, no. 8, pp. 739– 744, 2010.
[6] Martin Ester, Hans-Peter Kriegel, Jorg Sander and Xiaowei Xu, "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise," Proceedings of 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96), Portland, Oregon, 1996.
[7] Jiawei Han, MIcheline Kamber and Jian Pei, "Data Mining Concepts And Techniques" III edition 2012

Author Profile

Ruchi Rathor received the B.Tech degree in Computer Science And Engineering from Amity University in 2012 and perusing M.E in Computer Engineering from Dr D .Y Patil School Of Engineering, Savitribai Phule Pune University, India.

Pankaj Agarkar received the B.E and M.Tech degree in Computer science and engineering. Also appearing for PhD. Currently working as Assistant Professor of Computer Engineering Department in Dr. D.Y.Patil School of Engineering, Savitribai Phule Pune University, India