A SURVEY ON USAGE ANALYTICS FOR IMAGING DIAGNOSTIC EQUIPMENTS

MERIN SUSANNA VERGHESE, PARVATHI R

School of Computing Science and Engineering, VIT University, Chennai, Tamil Nadu, India. Email: parvathi.r@vit.ac.in

ABSTRACT

The main aim of the paper is to find the best predictive algorithm that can be used to predict the usage of the imaging diagnostic equipment in the near future. The growing number of patients are not able to receive the image diagnostic test within a short waiting time. The goal is to keep the waiting time low. The rising demand for the image diagnostics equipment has to be estimated to improve the service.

Keywords: Predictive analytics, Descriptive analytics, Machine learning algorithm, Imaging diagnostics, Regression.

INTRODUCTION

Over the past 30 years, imaging diagnostics have been transformed by developments in information technology and digital electronics than any other clinical service diagnostics. The number of patient suffering due to more waiting time for image diagnostic test shows a worrying upward trend. First, a descriptive analytic study has to be done on the present data, which will help to understand the details about the current scenario. Moreover, then a predictive analytics have to be done to predict the demand of the imaging diagnostic equipment to ensure quality and performance improvement. For this study, data of diagnostic examination of inpatient and outpatient are needed. The study should also include the number of people coming for image diagnostics to a hospital from that region and outside the region. The population growth will also play an important part in this prediction. Today's population is getting older, so there will be a need for more health care. The regional data and the population data have to be considered. The improved patient experience is the indication of the quality service offered. Timely access to the image diagnostic equipment services is very important. The installation and use of image diagnostic equipments are complicated and expensive, so the demand has to be predicted accurately. The treatment waiting time is also depending on the diagnostic test waiting time. To get the proper treatment faster, the diagnostic test has to be done faster. The statistics say that the number of people waiting is increasing year by year. This paper aims to deliver a suitable solution to reduce the waiting time. The intention is to develop a predictive analytic model, based on detailed data from inpatient and outpatient, and point out insights, on which hospitals should be installed which services and in what capacity.

METHODS FOR PREDICTION

Decision tree algorithm

The decision tree is a decision support data mining algorithm, which can be used for prediction. It is capable of handling continuous attribute. This can be used in regression as well as in association data mining. The decision tree is faster than neural network. For classification and prediction, a shorter learning curve is mainly used to represent knowledge [1]. The attributes will be classified down the tree to the appropriate leaf node from the root node by sorting. Moreover, the nodes are branched based on the condition specified using "If Then." The main challenge is to find the best attribute to divide the dataset.

There are several nodes in a decision tree. The nodes indicate the conditional attributes (s1, s2, and s3). Moreover, the branches show the value (v). The last node in the tree shows the decision (d1,d2) which is a binary value. A model of the decision tree is presented in Fig. 1 [1].

Artificial neural networks (ANN's) algorithm

The ANN is mainly used for prediction. It is being used for prediction despite its complexity and difficulty [2]. After learning from existing data or training data, ANNs capable of predicting new observations. ANN is biologically inspired analytical techniques. Complex non-linear functions can be modeled using the ANN. For the modeling, neurological functions of the brain are used [3]. Even if noise is present in the training data, ANN will find hidden and strong non-linear dependencies. Backpropagation, ART, and mark network are various types of neural network that can be used for prediction. This is mainly used for predicting time series Fig. 2 demonstrates an ANN [1].

Support vector machine (SVM)

SVM is a training, which is based on decision planes that segregate feature across the boundary of decisions. The point between the classes that form the decision boundary is called as support vector. SVM concerns itself with the problem of building consistent estimators from data.

K-nearest neighbor

K-nearest neighbor stores the complete training data. It is considered as a type of instance-based learning, which is also a simplest machine learning algorithm. Continuous variables can be estimated using the K-nearest neighbor regression. Inverse of their distance is used to calculate the weighted average in this algorithm [2]. The neighbors are selected from a set of objects. Those objects are correctly classified. K-nearest neighbor is having strong consistency results. The disadvantage of K-nearest neighbor is a computational cost which is expensive for large data set.

Naive Bayes algorithm

Naive Bayes algorithm is a predictive model, based on Bayes rule of conditional probability. This can predict the class of test dataset. It needs only a small amount of training data. In the structure of this model, there will be a node for each attribute. The input for the problem statement will be numerical, but the Naïve Bayes will perform well for categorical data. The directed edges interconnect these nodes to form directed acyclic graph. The Bayes theorem provides the below equation to calculate the posterior probability.

$$P(S|T) = \frac{P(T|S) * P(S)}{P(T)}$$

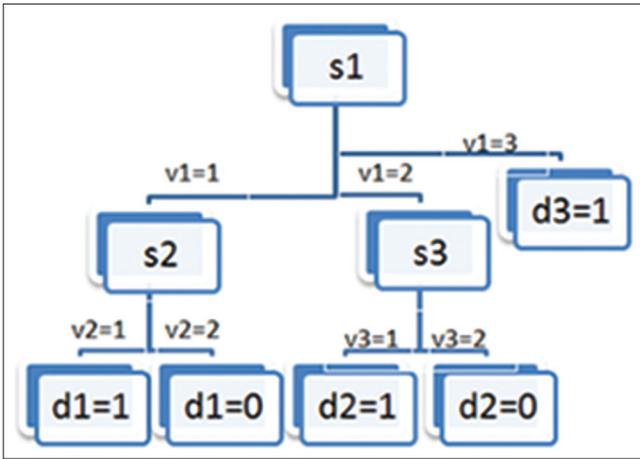Equation 1: Probability equation of naive Bayes
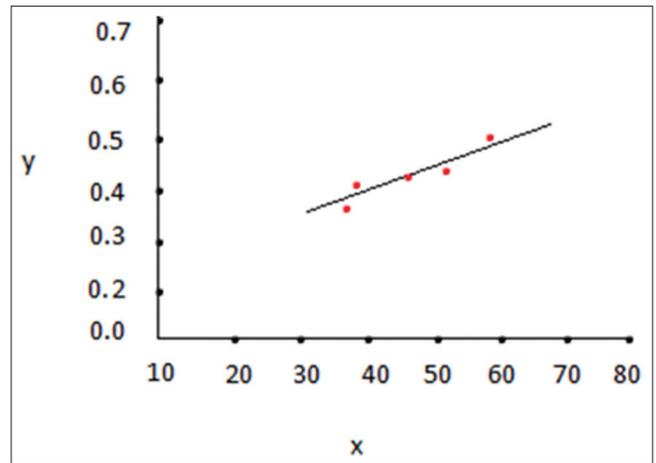
Fig. 1: Decision tree model
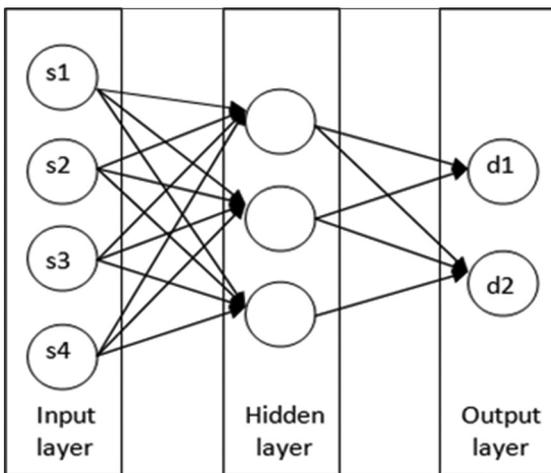

Fig. 3: Ordinary linear regression graph

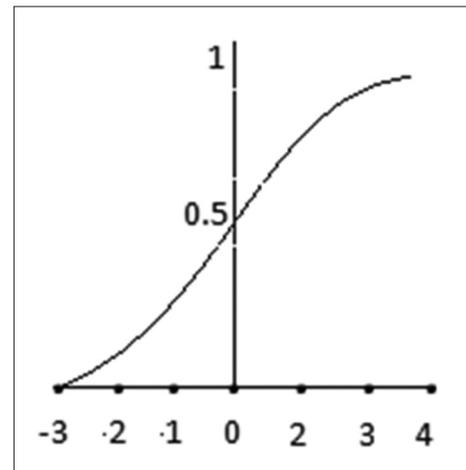Fig. 2: Artificial neural network model

Fig. 4: Standard logistic sigmoid function

Linear regression

A linear relationship is used to predict the numeric values of y for a given value of x using a straight line called as the regression line. Then, find the slope and the y-intercept of that regression line, then we can give a value for x and predict the value of y from x. Line of linear regression is using the equation of line that is y = a+bx, in which x and y are explanatory and dependent variable, respectively. The slope of the line is taken as "b" in the equation. In the equation, "a" denotes the intercept. This gives optimal result to the model with variables having a linear relationship (Fig. 3).

Logistic regression

Logistic regression is a predictive analysis model [4]. Categorical or binary-dependent variable is present in this. A logistic function is used to find the probability. Equation 2: Is logistic function. This in turn measures the relationship between dependent and independent variables. It is a generalized linear model. The equation is a logistic function or a logistic curve (Fig. 4).

$$f(x) = \frac{L}{1 + e^{-k(x-x_0)}}$$

Equation 2: Probability function

PROPOSED WORK

As an initial step, the data have to be taken care. The data of inpatient and outpatient imaging exams along with that equipment data, municipality data, and population data have to be considered. Gamma

camera, mammography with simple command, mammography with stereotactic, X-ray up to 100 mA, X-ray more than 500 mA, magnetic resonance imaging, and ultrasound are some of the devices that we are considering in this work.

Cleaning and filtering are being done for making the source data appropriate for simulation and data visualization, this is the second step in the workflow. For filtering and cleaning, data algorithm has to be developed and implemented. In the third part of the workflow, the data have to undergo classification, organization, aggregation, and transformation. Other parameters such as age, sex, and population should also to be considered.

The fourth step in the workflow is descriptive analytics. As the name implies, the descriptive analytics will describe or summarize the data. It provides insight to the past by giving answer to the question "what happened?" Data visualization tools such as Tableau can be used to get a better understanding of the data. For this descriptive study, we should be able to get the details of the image diagnostic equipments utilization in the hospitals. We can also look at the number of people coming from other regions to a hospital for image diagnostics. A complete understanding of the data will help us to generate the appropriate and accurate output.

The fifth step is predictive analytics. It will predict the possible outcome based on the input data. R programming can be used for modeling purposes. Then, the waiting time has to be estimated. On that optimization has to be done to predict the appropriate demand

for supply. The final step is to represent the outcome of the model in a simple and attractive way. For this also, dashboards and data visualization using R or Tableau can be used. Fig. 5 shows the workflow diagram of the proposed work. The expected outcome is the number of the imaging diagnostics that have to be supplied, respectively, to the estimated population.

As discussed before, the main aim is to minimize the time of waiting in hospitals for imaging diagnostics. Queuing theory-based modeling is suggested for determining the waiting time. With the help of the queuing theory, the waiting time can be predicted. This gives the information that in which hospital the patients have to wait for long to get imaging diagnostics. This indicates that particular hospital has to be installed with more number of machines to reduce the waiting time.

Markov chain is a theorem in queuing theory that helps to reduce the waiting time. There are different types of queuing system in queuing theory. M/M/C/K, this can be used in our project for calculation and prediction of waiting time. In this, the first M indicates the Poisson arrival distribution. The second M is for exponential distribution of service time. C stands for the number of servers in the system. K gives the count of customers that can accommodate the queue. Fig. 6 illustrates a simple queuing theory system.

LITERATURE SURVEY

The healthcare data like patient's details can be stored digitally and how useful information can be generated out of it using analytics techniques and tools were discussed in [5]. Three types of simulation modeling in medical diagnosing [6,7]. In [3] discusses different data mining algorithms are being discussed and a comparative study was done on the accuracy of the algorithm. The insights about simulators
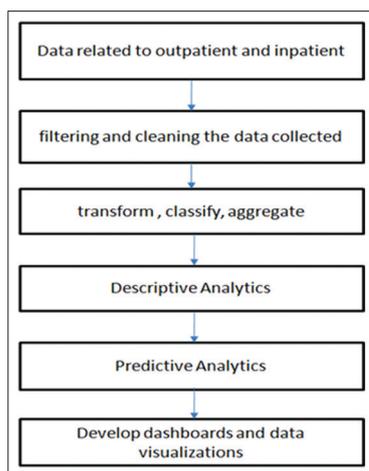

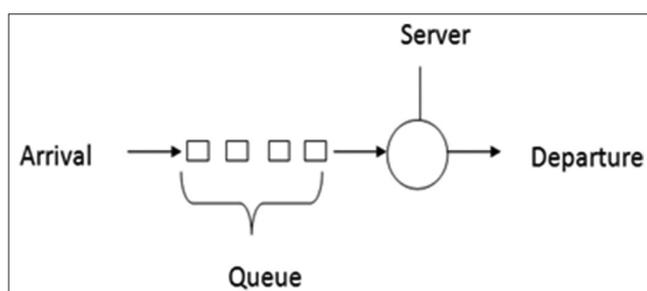**Fig. 5: Workflow diagram**

Fig. 6: Simple queuing system

that can be used to decrease the waiting time at a medical imaging center[8].

A survey of machine learning techniques such as decision tree, support vector machine, Nearest Neighbor, etc., that can be used for data mining is being discussed in [2]. The policies for maximum waiting time standards which is followed in the hospital is being described in [9]. Logistic regression and artificial neural network were discussed in [10]. The logistic regression approach looks good for this work. In [11], the information about waiting time in medical imaging department. the different types of diagnostic imaging and their benefits were discussed [12]., In a new Markov model were discussed [13]. A Queue Selector that can be used for the improvement of Utilization of the server [14].How queuing theory can be applied in bank sector to optimize the service is being discussed in [15]. M/Mc model is the discussion topic in the paper [16,17]

CONCLUSION

Diagnostic imaging contributes a major part in planning for treatment. Through different perspectives, the variation in diagnostic imaging services has to be analyzed and the predictive model has to be applied to supply the equipments according to the future demand. The demand for the equipments depends on the population growth and the diseases. By considering the dependencies between the variables, the linear regression will be a suitable algorithm for prediction. The aim is to ensure that the services are planned and available to accommodate the increase in demand for image diagnostic equipments. Effective utilization of the image diagnostic equipments will help to save cost also. Thus, the quality of service received by the people will be better. The earlier image diagnosis of disease through imaging diagnostic equipment will help the patient to receive the treatment earlier, and the chance of getting cured will be high. Hence, that we can assure a safe and healthier future. By this prediction, the number of equipments can be made to a necessary count that helps people to get a better treatment at the right time.

REFERENCES

1. Magsino ER, Ho IH. An intelligent Highway Tollgate Queue Selector for Improving Server Utilization and Vehicle Waiting Time. Region 10 Symposium IEEE; 2016.
2. Jansen FJ. Aggregate Simulation Modeling of an Mri Department using Effective. Proceedings of the Winter Simulation Conference; 2012.
3. Dreiseitl S. Logistic regression and artificial neural network classification models: A methodology review. J Biomed Inf 2002;35:352-9.
4. Amooee G. A comparison between data mining prediction algorithmsfor fault detection. Int J Comput Sci 2011;8(6):425-31.
5. Akbari H, Fernando X. A Queuing Model for Health Care Pharmacy Using Software Arena. International Conference on Industrial Engineering and Operations Management; 2015.
6. Xiao H, Zhang G. The Queuing Theory Application in Bank Service Optimization. IEEE; 2010.
7. Ojha M, Mathur K. Proposed Application of Big Data Analytics in healthcare at Maharaja Yeshwantrao Hospital. 3rd MEC International Conference on Big Data and Smart City; 2016.
8. Fotiadou A. Choosing and Visualizing Waiting Time Indicators in Diagnostic Medical Imaging Department for Different Purposes and Audiences. Master's Programme in Health Informatics Spring Semeste; 2013.
9. Osofisan A. Evaluation of Predictive Data Mining Algorithms in Erythemato-Squamous Disease Diagnosis. Department of Computer Science, ModibboAdama University of Technology.
10. Mapp G, Thakker D, Germikonakli O. Exploring a new Markov Chain Model for Multiqueue System. 12th International Conference on Computer Modeling and Simulation; 2010.
11. Maritza L. A Comparison of Several Predictive Algorithms for Collaborative Filtering on MultiValuedRatings. ACM Symposium on Applied Computing; 2004.
12. Nuti S, Vainieri M. Managing waiting times in diagnosticmedical imaging. BMJ Open 2012;2(6):e001255.

13. Parkin E. NHS Maximum Waiting Times Standards and Patient Choice Policies. UK House of Common Library; 2016.
14. Palmer PE. Diagnostic Imaging in the Community. A Manual for Clinics and Small Hospitals; 2011.
15. Ramis FJ. A Simulator to Improve Waiting Times at a Medical Imaging Center. Proceedings of the Winter Simulation Conference; 2008.
16. Sharma S. Machine Learning Techniques for Data Mining: A Survey. IEEE International Conference on Computational Intelligence and Computing Research; 2013.
17. Wang Y, Qian C, Cao J. Optimized M/M/C Model and Simulation for Bank Queuing System. IEEE International Confernce on Software Engineering and Service Sciences; 2010.