ABSTRACT

Hospital recommendation System is considered as an important factor in health care sector for managing the administrative, financial and clinical aspects of a hospital. Due to data mining progress in biomedical and healthcare communities, accurate study of medical data benefits early disease recognition, patient care and community services. When the quality of medical data is incomplete the exactness of study is reduced. Moreover, different regions exhibit unique appearances of certain regional diseases, which may results in weakening the prediction of disease outbreaks. In the proposed system, it provides machine learning algorithms for effective prediction of various disease occurrences in disease-frequent societies. It experiment the altered estimate models over real-life hospital data collected. To overcome the difficulty of incomplete data, it use a latent factor model to rebuild the missing data. It experiment on a regional chronic illness of cerebral infarction. Using structured and unstructured data from hospital it use Machine Learning Decision Tree algorithm. It predicts probable diseases and hospitals by mining data sets. To the best of our knowledge in the area of medical big data analytics none of the existing work focused on both data types. Compared to several typical estimate algorithms, the calculation exactness of our proposed algorithm reaches 94.8% with a convergence speed which is faster than that of the Decision tree disease risk prediction algorithm.

Table of contents

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

1.1 OVERVIEW

According to the statistics of National Health and Family Planning Commission, there are a total of 991,632 medical institutions in China until Nov. 2016 where new institutions are increased from Nov. 2015. Nevertheless, the difficulty of getting medical care remains as one of China's major livelihood issues. A survey on Peking University First Hospital, which is one of the most famous hospitals in Beijing, indicates that more than 45% outpatients have to wait for over two hours after the registration, whereas 85% of them have less than 10 minutes for the doctor's inquiry. This phenomenon is actually common in China's 776 Top-Class hospitals1. The reason is that many hospitals lack publicity and are not familiar to most patients. Having no way of knowing whether there is a good enough and less crowded hospital nearby, people have no choice but go to those congested famous hospitals regardless of the severity of the disease. In fact, most outpatients with a mild disease may expect a quick treatment yet have a low requirement for the hospital's treatment ability. It is imperative to find a simple and convenient way to access the crowd status and basic information of the neighboring hospitals. There are several standard ways for crowd counting, such as approaches based on video or beacon. However, these methods rely on surveillance data or wireless network data, and any company or non-governmental organization can hardly gather these data of all hospitals even in one city, not to mention in any larger range. Fortunately, location-based service (LBS) big data offers a potential solution to this dilemma. LBS data have two unique properties: a) LBS data naturally belongs to the service providers. Using it for population density estimation does not involve asking hospitals for any help. b) LBS data is sufficient for population density estimation with its copiousness and vast area coverage. As smartphone is popular, several main LBS providers in China have hundreds of millions of users and preserve billions LBS request logs every day. For instance, Baidu Map has over 300 million active users and gets 23 billion LBS request logs in China each day on average. Taking advantage of these two properties, we present a novel (nearly) real-time model for counting and predicting the crowd of all hospitals in the city based on LBS big data. What's more, the distribution of people's residence time can be figured out with the LBS data. Relying on the location logs of Baidu LBS and Baidu Points of Interests data, we have designed an APP2 to provide users with a) hospitals' real-time and predicted outpatient density, b) official level,

c) distance from the user to each hospital and the corresponding path planning.

To the best of our knowledge, it is the first hospital recommendation system based on the population density analysis with LBS big data. Note that Baidu LBS data has incorporated the information of GPS, WiFi, and Cellular network, so it is a relatively high-quality data source. Since the LBS data we use is location logs of mobile APPs, we will use data and logs interchangeably to denote the LBS data. Our contributions can be summarized into three innovations conquering three main challenges of this work. The first innovation is the model for detecting the types of people around hospitals to address highly noisy LBS data. Although our data source is relatively high-quality, it is still temporally.

1.2 SCOPE AND OBJECTIVE

The proposed software product is the hospital recommendation management system the system will be used in any hospital dispensary or pathology in any of the near location for hospital we have choose that one the current system in use paper based system it is slow and cannot provide update lists of patients within a reasonable timeframe.

The project of the hospital recommendation system is aimed to develop to maintain the day to day state of admission to computerize all details regarding patient details in this we have near location any hospital.

1.3 DOMAIN OVERVIEW

Machine Learning is the most popular technique of predicting the future or classifying information to help people in making necessary decisions. Machine Learning algorithms are trained over instances or examples through which they learn from past experiences and also analyze the historical data. Therefore, as it trains over the examples, again and again, it is able to identify patterns in order to make predictions about the future. Data is the core backbone of machine learning algorithms. With the help of the historical data, we are able to create more data by training these machine learning algorithms. For example, Generative Adversarial Networks are an advanced concept of Machine Learning that learns from the historical images through which they are capable of generating more images. This is also applied towards speech and text synthesis. Therefore, Machine Learning has opened up a vast potential for data science applications.

1.3.1 MACHINE LEARNING

Machine Learning combines computer science, mathematics, and statistics. Statistics is essential for drawing inferences from the data. Mathematics is useful for developing machine learning models and finally, computer science is used for implementing algorithms. However, simply building models is not enough. You must also optimize and tune the model appropriately so that it provides you with accurate results. Optimization techniques involve tuning the hyper parameters to reach an optimum result.

The world today is evolving and so are the needs and requirements of people. Furthermore, we are witnessing a fourth industrial revolution of data. In order to derive meaningful insights from this data and learn from the way in which people and the system interface with the data, we need computational algorithms that can churn the data and provide us with results that would benefit us in various ways. Machine Learning has revolutionized industries like medicine, healthcare, manufacturing, banking, and several other industries. Therefore, Machine Learning has become an essential part of modern industry. Data is expanding exponentially and in order to harness the power of this data, added by the massive increase in computation power, Machine Learning has added another dimension to the way we perceive information. Machine Learning is being utilized everywhere. The electronic devices you use, the applications that are part of your everyday life are powered by powerful machine learning algorithms.

With an exponential increase in data, there is a need for having a system that can handle this massive load of data. Machine Learning models like Deep Learning allow the vast majority of data to be handled with an accurate generation of predictions. Machine Learning has revolutionized the way we perceive information and the various insights we can gain out of it. These machine learning algorithms use the patterns contained in the training data to perform classification and future predictions.

Whenever any new input is introduced to the ML model, it applies its learned patterns over the new data to make future predictions. Based on the final accuracy, one can optimize their models using various standardized approaches. In this way, Machine Learning model learns to adapt to new examples and produce better results.

Types of Machine Learning

Machine Learning Algorithms can be classified into 3 types as follows –

1. Supervised Learning
2. Unsupervised Learning
3. Reinforcement Learning

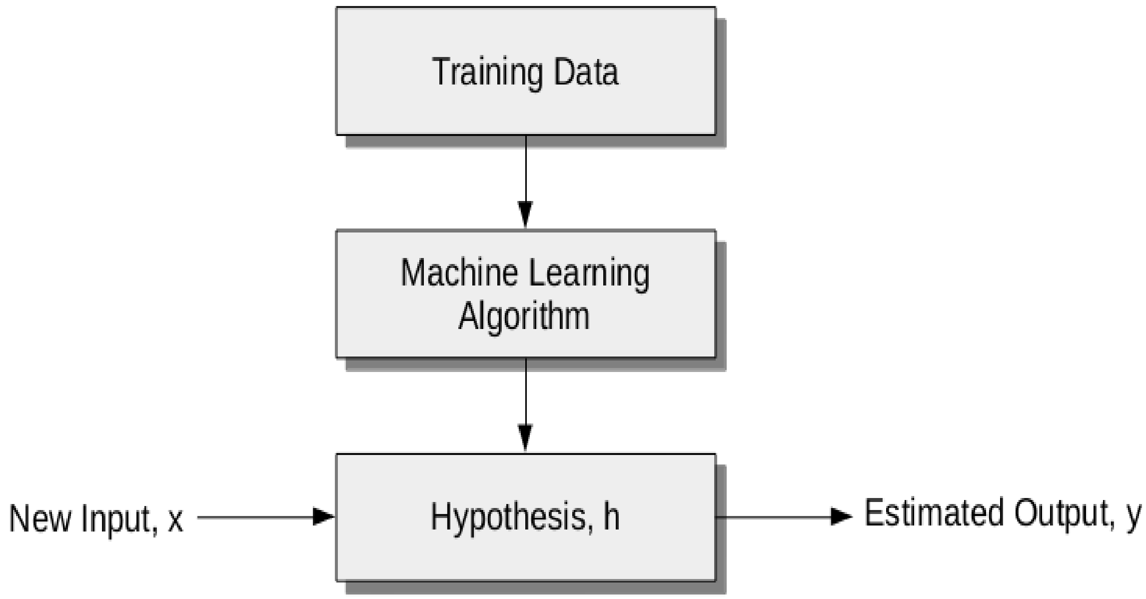

Fig -1.1 Supervised Architecture

SUPERVISED LEARNING

In the majority of supervised learning applications, the ultimate goal is to develop a finely tuned predictor function h(x) (sometimes called the “hypothesis”). “Learning” consists of using sophisticated mathematical algorithms to optimize this function so that, given input data x about a certain domain (say, square footage of a house), it will accurately predict some interesting value h(x) (say, market price for said house).

$$h(x_1, x_2, x_3, x_4) = \theta_0 + \theta_1 x_1 + \theta_2 x_3^2 + \theta_3 x_3 x_4 + \theta_4 x_1^3 x_2^2 + \theta_5 x_2 x_3^4 x_4^2$$

This function takes input in four dimensions and has a variety of polynomial terms. Deriving a normal equation for this function is a significant challenge. Many modern machine learning problems take thousands or even millions of dimensions of data to build predictions using hundreds of coefficients. Predicting how an organism’s genome will be expressed, or what the climate will be like in fifty years, are examples of such complex problems.

Under supervised ML, two major subcategories are:

- Regression machine learning systems: Systems where the value being predicted falls somewhere on a continuous spectrum.
- Classification machine learning systems: Systems where we seek a yes-or-no prediction.

In practice, x almost always represents multiple data points. So, for example, a housing price predictor might take not only square-footage (x1) but also number of bedrooms (x2), number of bathrooms (x3), number of floors (x4), year built (x5), zip code (x6), and so forth. Determining which inputs to use is an important part of ML design. However, for the sake of explanation, it is easiest to assume a single input value is used.

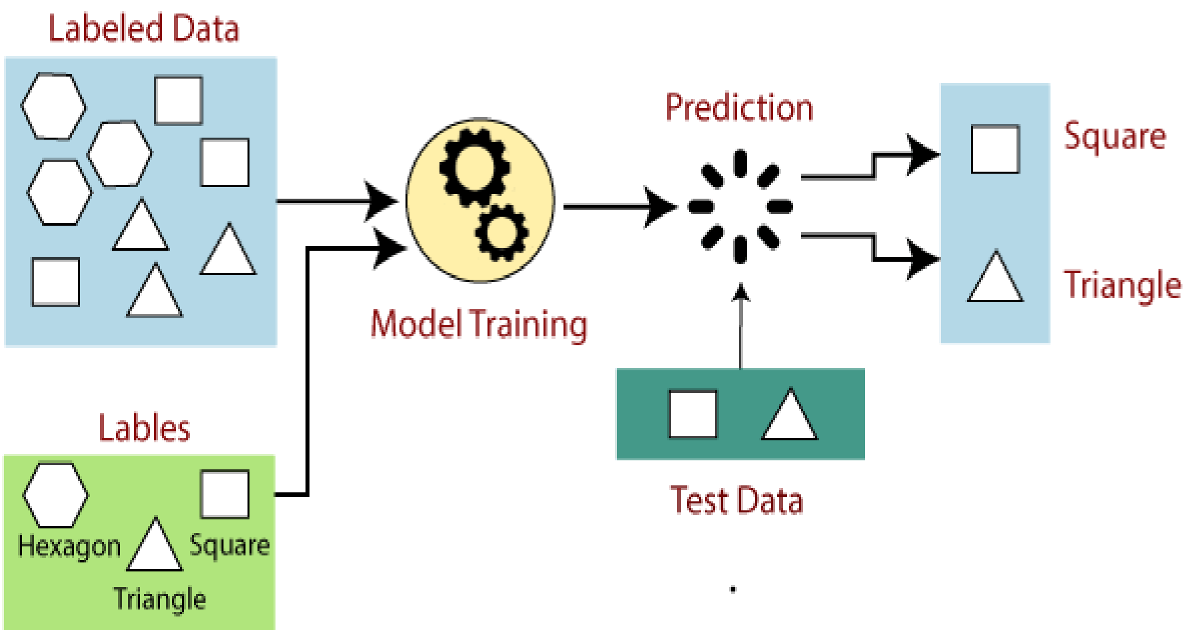

Fig-1.2 Block Diagram

Steps Involved in Supervised Learning:

- First Determine the type of training dataset
- Collect/Gather the labelled training data.
- Split the training dataset into training dataset, test dataset, and validation dataset.
- Determine the input features of the training dataset, which should have enough knowledge so that the model can accurately predict the output.
- Determine the suitable algorithm for the model, such as support vector machine, decision tree, etc.
- Execute the algorithm on the training dataset. Sometimes we need validation sets as the control parameters, which are the subset of training datasets.
- Evaluate the accuracy of the model by providing the test set.

REGRESSION

Regression algorithms are used if there is a relationship between the input variable and the output variable. It is used for the prediction of continuous variables, such as Weather forecasting, Market Trends, etc.

1. Linear Regression
2. Regression Trees
3. Non-Linear Regression
4. Bayesian Linear Regression
5. Polynomial Regression

CLASSIFICATION

Classification algorithms are used when the output variable is categorical, which means there are two classes such as Yes-No, Male-Female, True-false, etc.

Spam Filtering,

- Random Forest
- Decision Tree
- Logistic Regression
- Support vector Machines

1.3.2 PROPOSED ALGORITHMS

Decision Tree Classification Algorithm

- Decision Tree is a supervised learning technique that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems. It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome.
- In a Decision tree, there are two nodes, which are the Decision Node and Leaf Node. Decision nodes are used to make any decision and have multiple branches, whereas Leaf nodes are the output of those decisions and do not contain any further branches.
- The decisions or the test are performed on the basis of features of the given dataset.