International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2018 IJSRCSEIT | Volume 3 | Issue 3 | ISSN : 2456-3307

Patient Centric Healthcare Analysis and Classification to Recommend Medicines

S. Pradeepkumar, S. Sabeek Ahamed, K. Ponnivalavan, B. Srivathsan,
Computer Science and Engineering Department, Saranathan College of Engineering, Trichy, Tamilnadu, India

ABSTRACT

Handling Bigdata is one of the major issues in the field of Medicine. The huge numbers of tablets, manufacturers, distributors, doctors and diseases leads a world with medical errorness and correptiveness. Our intention is to reduce the fraudulent and the deaths rate due to this. By analyzing patients data and tablets data we conclude that the perfect solution for a patient by make use of machine learning techniques. By implementing our project every patient, doctor and medicines prescribed are monitored and every patient assure that they are treated by correct treatment.

Keywords : Decision Tree, Machine Learning, Medical Diagnosis, Random Forest, Medical Imaging

I. INTRODUCTION

Data science plays an important role in many industries. In facing massive amount of heterogeneous data, scalable machine learning and data mining algorithms and system become extremely important for data scientist. The growth of volume, complexity and speed in data drives the need for scalable data analytic algorithms and systems.

In healthcare, large amount of heterogeneous medical data have become available in various healthcare organizations. This data could be an enabling resource for driving insights for improving care delivery and reducing waste. The enormity and complexity of these datasets present great challenges in analysis and subsequent applications to a practical clinical environment. Machine learning algorithms were from the very beginning designed and used to analyze medical data sets. Today machine learning provides several indispensable tools for intelligent data analysis. Especially in the last few years, the digital revolution provided relatively inexpensive and available means to collect and store the data. Modern hospitals are well equipped with monitoring and other data collection devices, and data is gathered and shared in large information systems. Machine learning technology is currently well suited for analyzing medical data, and in particular there is a lot of work done in medical diagnosis in small specialized diagnostic problems. Data about correct diagnoses are often available in the form of medical records in specialized hospitals or their departments. All that has to be done is to input the patient records with known correct diagnosis into a computer program to run a learning algorithm. This is of course an over simplification, but in principle, the medical diagnostic knowledge can be automatically derived from the description of cases solved in the past. The derived classifier can then be used either to assist the physician when diagnosing new patients in order to improve the diagnostic speed, accuracy and/or reliability, or to train students or physicians non-specialists to diagnose patients in a special diagnostic problem. The aim of this paper is to provide an overview of the development of the intelligent data analysis in medicine from a machine learning perspective a historical view, a state of the art view and a view on some future trends in this subfield of applied artificial intelligence, which are respectively described in the following three sections. None of the three sections is intended to provide a comprehensive overview but rather describe some subareas and directions which from my personal point of view seem to be important for medical diagnosis. In the historical overview We emphasize the Random forest and decision trees. One or two representatives from each branch of machine learning, when applied to several medical diagnostic task. The future trends are illustrated by two case studies. I describe a recently developed method for dealing with reliability of decisions of classifiers, which seems to be promising for intelligent data analysis in medicine and an approach to using machine learning in order to verify some unexplained phenomena from complementary medicine, which is not (yet) approved by the orthodox medical community but could in the future play an important role in overall medical diagnosis and treatment.

II. HISTORICAL OVERVIEW

As soon as electronic computers came into use in the fifties and sixties, the algorithms were developed that enabled modeling and analyzing large sets of data. In medical field the doctor having the full responsibility to prescribe and treat the patient. This may leads almost good for all peoples even though peoples died because of medical errors.

Instead of using doctor’s past knowledge for recommending medicines an machine learning algorithm will be a replacement with 100% results in patients survivability.

From the very beginning three major branches of machine learning emerged. Classical work in symbolic learning is described by Hunt et al. (1966), in statistical methods by Nilsson (1965) and in neural networks by Rosenblatt (1962). Through the years all three branches developed advanced methods (Michie et al., 1994): statistical or pattern recognition methods, such as the k-nearest neighbors, discriminant analysis, and Bayesian classifiers, inductive learning of symbolic rules, such as top-down induction of decision trees, decision rules and induction of logic programs, and artificial neural networks, such as the multilayered feed forward neural network with back propagation learning, the Kohonen’s self-organizing network and the Hopfield’s associative memory.

III. BACKGROUND DETAILS

Classification algorithms are widely used in various medical applications. Data classification is a two phase process in which first step is the training phase where the classifier algorithm builds classifier with the training set of tuples and the second phase is classification phase where the model is used for classification and its performance is analyzed with the testing set of tuples. Brief about the various classification algorithms in medical domain are:

IV. DECISION TREE ALGORITHM

The decision tree is one of the classification algorithms. The learning algorithm applies a divide and-conquer strategy to construct the tree. The sets of instances are associated by a set of attributes. A decision tree comprises of nodes and leaves, where nodes represent a test on the values of an attribute and leaves represent the class of an instance that satisfies the conditions. The outcome is „true‟ or „false‟. Rules can be derived from the path starting from the root node to the leaf and utilizing the nodes along the way as preconditions for the rule, to predict the class at the leaf. The tree pruning has to be carried out to remove unnecessary preconditions and duplications.

One best medical plan among lots of TB treatments is identified by applying decision tree algorithm in past datasets. It deals like the way explained above.

V. RANDOM FOREST

Random forest algorithm is one of the best among classification algorithms and is able to classify large amounts of data with high accuracy. It is an ensemble learning method building models that constructs a number of decision trees at training time and outputs the modal class out of the classes predicted by individual trees. It is a combination of tree predictors where each tree depends on the values of a random vector sampled independently with the same distribution for all the trees in the forest. The basic principle is that a group of "weak learners" can come together to form a "strong learner".

The Random forest is used to identify whether a patient with TB symptoms is having the possibility of TB disease. It recommends the patient to either to take the TB tests or directly tells to the patient you are not having the TB. Random forest algorithm applied in three places to recommending medicines and every time interval of two months.

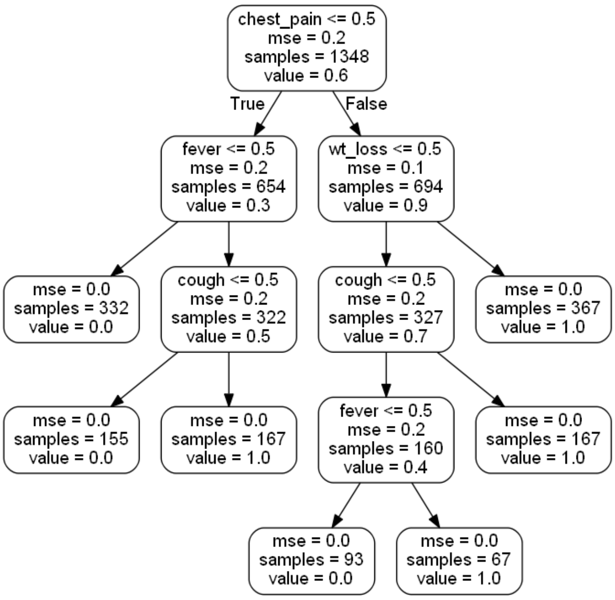

Figure 1

This figure shows the basic implementation of random forest algorithm applied in first phase. It tells whether a patient with TB disease need to test or not.

SPECIFICREQUIREMENTS FOR MACHINE LEARNING

For a machine learning (ML) system to be useful in solving medical diagnostic tasks, the following features are desired: good performance, the ability to appropriately deal with missing data and with noisy data (errors in data), the transparency of diagnostic knowledge, the ability to explain decisions, and the ability of the algorithm to reduce the number of tests necessary to obtain reliable diagnosis.

VI. GOOD PERFOMANCE

The algorithm has to be able to extract significant information from the available data. The diagnostic accuracy on new cases has to be as high as possible. Typically, most of the algorithms perform at least as well as the physicians and often the classification accuracy of machine classifiers is better than that of physicians when using the same description of the patients. Therefore, if there is a possibility to measure the accuracy of physicians, their performance can be used as the lower bound on the required accuracy of a ML system in the given problem. In the majority of learning problems, various approaches typically achieve similar performance in terms of the classification accuracy although in some cases some algorithms may perform significantly better than the others (Michie et al., 1994). Therefore, Apriori algorithm.

Almost none of the algorithms can be excluded with respect to the performance criterion. Rather, several learning approaches should be tested on the available data and the one or few with best estimated performance should be considered for the development of the application.

The ares we need to taking consideration is more so a doctor with good in one field is not recommended to treat a patient nowadays. This also leads a patient treated with wrong medicine . More than a doctor a machine learning artificial algorithm is needed in medical field.

VII. DEALING WITH MISSING DATA

In medical diagnosis very often the description of patients in patient records lacks certain data. ML algorithms have to be able to appropriately deal with such incomplete descriptions of patients.

VIII. DEALING WITH NOISY DATA

Medical data typically super from uncertainty and errors. Therefore machine learning algorithms appropriate for medical applications have to have elective means for handling noisy data.

TRANSPERANCY OF DIAGNOSTIC KNOWLEDGE

The generated knowledge and the explanation of decisions should be transparent to the physician. She should be able to analyze and understand the generated knowledge. Ideally, the automatically generated knowledge will provide to the physician a novel point of view on the given problem, and may reveal new interrelations and regularities that physicians did not see before in an explicit form.

IX. EXPLANATIONABILITY

The system must be able to explain decisions when diagnosing new patients. When faced with an unexpected solution to a new problem, the physician shall require further explanation , otherwise she will not seriously consider the system's suggestions. The only possibility for physicians to accept a "black box" classifier is in the situation where such a classifier outperforms by a very large margin all other classifiers including the physicians themselves in terms of the classification accuracy. However, such situation is typically highly improbable.

REDUCTION OF THE NUMBER OF TESTS

In medical practice, the collection of patient data is often expensive, time consuming, and harmful for the patients. Therefore, it is desirable to have a classifier that is able to reliably diagnose with a small amount of data about the patients. This can be verified by providing all candidate algorithms with a limited amount of data. However, the process of determining the right subset of data may be time consuming as it is essentially a combinatorial problem. Some ML systems are themselves able to select an appropriate subset of attributes, i.e., the selection is done during the learning process and may be more appropriate than others that lack this facility.

Mathematics behind Random Forests and decision tree

X. RANDOM FOREST

Random forest has more accuracy than the single-tree model, and handles a very large number of input variables. Besides, it provides an experimental way to detect variable interactions, etc. Instead of using all training data, a random sample of N observations with replacement is chosen to build a tree. In the tree building process, for each node, a random subset of the predictor variables is considered as possible splitters for each node, a predictor excluded from one split is allowed to be used as splitters in the same tree. Repeat the above procedure until a large number of trees are constructed. The average of the predicted value in regression trees are computed as the predicted value and the most frequently predicted category in the classification trees are considered to be the predicted category.

1) **Theorem 1 (Chebyshev inequality)**

If is a random X variable withσ standard deviation and μ mean , then for any $\varepsilon > 0$

$$P((|X-\mu|) > \epsilon) \leq \frac{\sigma^2}{\epsilon^2}$$

2) **Theorem 2 (Bounded convergence theorem)**

Given a sequence

$h_1(X), h_2(X) \ldots, h_n(X)$ with $h_1(X) \leq M$ for M>0 defined on a space S of finite measure then

$$\lim_{k\to\infty}\int dX\, h_k(X) \to \int dX \lim_{k\to\infty} h_k(X)$$

Data set contains the different symptoms of the patients tested for TB. Patients tested for TB positive and negative are analyzed with the entered symptoms that the patient have.

Each symptom with test result are taken as analyses in every modules as a separate decision tree. Every features attribute added to the resultant decision tree again analyzed finally a decision is made by applying all features above the dataset. The outcome of this phase is tell the patient he patient to test for TB or not. Random forest again used to find the medicine for the patient with the TB test results are inputted to the model of random forest with dataset contains tablets details and past history of the patient. In this phase random forest result is the medicine with particular manufacturer.

XI. DECESION TREE

The difference between the random forest and decision tree is random forest is the combinations of decision trees. Applying again and again the decision tree over the result of it gives the random forest algorithm. A decision made by make use of the dataset we have and input from the patient analyzed and output made. The output of decision tree is plan of the TB patient treatment.

Out of four plan one plan is picked up by decision tree algorithm based on the input that is the symptoms and result of the TB tests. The total number of tablets and manufacturer that a doctor prescribed for and each patient medical history monitored.

XII. CONCLUSION

A Successful implementation of machine learning algorithms in medical diagnosis can help the integration of computer based systems in the healthcare environment. Especially in developing and highly populated country like India where mortality ratio is high and there is only one doctor for every 1700 persons, machine learning techniques in medical diagnosis can assist physicians to diagnose and cure diseases at early stage. Technology can no way replace a doctor’s experience and expertise, but it can take care of relatively straightforward yet time consuming diagnostic tasks and doctors can take up clinically more demanding procedure.

The dependence on medical images for diagnosing a disease is on rise. Since interpreting modern medical images is becoming increasingly complex, machine learning algorithms in medical imaging can provide significant assistance in medical diagnosis. They can help interns or less experienced physicians to reliably evaluate medical images and thus improve their diagnostic accuracy, sensitivity and specificity. Protein function prediction is another important area where machine learning techniques have a vital role to play. Machine learning techniques could be used for large scale and complex biological data analysis as these techniques are efficient and inexpensive in solving bioinformatics problems. The research in this area will not only be beneficial for physicians in terms of diagnosing diseases, it may also help health planners for diagnosing and preventing diseases at a large scale.

XIII. REFERENCES

[1]. J.S.Saleema, N.Bhagawathi, S.Monica, P.DeepaShenoy, K.R.VenugopalandL.M.Patnaik," Cancer Prognosis Prediction using Balanced Stratified Sampling" International Journal on Soft Computing, Artificial Intelligence and Applications (IJSCAI), Vol.3, No. 1, February 2014..

[2]. K.Senthamaraikannan, N. Senthilvel Murugan, V. Vallinayagam and T. Viveka, "Analysis of Liver Cancer DNA Sequence Data using Data Mining " International Journal of Computer Applications (0975 – 8887) Volume 61– No.3, January 2013

[3]. Anju Jain Asst. Professor, Deptt of CSE, G.J.U. S&T, Hisar, Haryana (India),"Machine learning techniques for medical diagnosis",University of delhi conference centre,27-sep-2015,www.conferenceworld.in,www.icstmdu.com(978-81-931039-6-8).

[4]. P.Dhivyapriya, Dr.S.Sivakumar Research Scholar , Assistant professor Department of Computer Science, Department of Computer Applications Thanthai Hans Roever College, Perambalur Tamil Nadu – India,"Classification of Cancer Dataset in Data Mining Algorithms Using R Tool"," International Journal on Soft Computing, Artificial Intelligence and Applications (IJSCAI), Vol.3, No. 1, February 2014..