Prediction &Validation of Heart/Cancer Diseases Using CART and ANN

BITS ZG628T: Dissertation

By S.B. Sivaraman - 2015HV92526

Dissertation work carried out at

Verizon Data Services India Pvt. Ltd - Chennai

BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE
PILANI (RAJASTHAN)

August 2017

Prediction &Validation of Heart/Cancer Diseases Using CART and ANN

BITS ZG628T: Dissertation

By S.B. Sivaraman - 2015HV92526

Dissertation work carried out at

Verizon Data Services India Pvt. Ltd - Chennai

Under the Supervision of

Mr.Thiru – Senior Architect, Verizon Data Services India Ltd,Chennai.

BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE

PILANI (RAJASTHAN)

August 2017

Outline of Presentation

1. Introduction
2. Issues & Challenges
3. Objective
4. Related Work
5. Problem Statement
6. Methodology
7. Literature References
8. Step by Step processing of the proposed work
9. Working Schedule

Introduction

- In healthcare and medical sector, there is need to provide accurate and precise diagnosis and treatment to patients in order to meet their requirements and provide them high quality and affordable care. Literature suggests that the quality of the service provided to the patients in healthcare organization refers to accurate diagnosis and providing them treatments that are instrumental in treating it efficiently.

- Furthermore, hospitals and healthcare organizations also focus on reducing their costs and therefore, they aim at reducing costs associated with clinical testing by using computer based systems and decision support systems.

- Majority of the hospitals have adopted hospital information systems to store, record and manage patient data.

Introduction Contn..

- These systems have large amount of data, which is stored in the form of images, text, charts and numbers. Furthermore, such data has hidden information, which can be extracted from the huge whirlpool of information to help clinical decision making.

- Data mining techniques are used effectively to improve the efficiency of classification and prediction systems and therefore, can aid medical practitioners in their decision making process.

- This can be beneficial in improving the quality of care for patients, while it can reduce operational costs and thus, can lay out the foundation for further clinical studies.

Issues & Challenges

1. DM tools have different capabilities and the user must be knowledgeable about these capabilities and differences.

2. The inability to choose the right tools for the appropriate tasks means misuse and possible failure: obtaining no answer, obtaining the wrong answer (even with right tools used) and misinterpreting the answer.

3. Another facet of technical difficulties is associated with creating and mining a data warehouse with data from different organizations.

4. Social concerns about the privacy of individuals, especially within health-care organizations – as the majority of data assessed through these organizations will contain pertinent health information.

Issues & Challenges Contn…

5. One of the biggest problems in data mining in medicine is that the raw medical data is voluminous and heterogeneous.

6. How to develop efficient algorithms for comparing content of two knowledge versions (before and after). This challenge demands development of efficient algorithms and data structures for evaluation of knowledge integrity in the data set.

7. How to develop algorithms for evaluating the influence of particular data modifications on statistical importance of individual patterns that are collected with the help of common classes of data mining algorithm.

Objective

- The goal of this research is to investigate how data can be extracted into useful information and can help healthcare practitioners in their decision making.

- The use of efficient and effective prediction system designed specifically for cancer and heart disease can be instrumental in improving the effectiveness of the healthcare organizations and can aid clinicians to take strong and effective decisions.

- It is essential that clinicians and patients are aware of the dangers of fatal diseases such as cancer and heart diseases and therefore, require efficient treatment. Consequently, modern healthcare organizations have adopted data mining techniques.

Objective Contn…

- In our work we will be using the dataset on heart diseases and cancer, which is available to the public by using CART (Classification and Regression Tree) for the data set management and for decision making and for validating the decision made Artificial Neural Network.

Related Work

- *Oleg Yu. Atkov et al. (Elsevier, 2012"Coronary heart disease diagnosis by artificial neural networks including genetic polymorphisms and clinical parameters" by)*, which is based on Artificial Neural Network and some of the clinical parameters.

- Accuracy of models was improved by a genetic algorithm with different optimization parameters including number of neurons in the hidden layer, number of inputs to the neural network, and slope coefficient of activation functions.

- Neuro-Solutions 5.0 development environment(Neuro-Dimension Inc., Gainesville, FL, USA) was used to check the possibilities of optimization.

Related Work Contn..

- In this work the major disadvantage is that the ANN inputs are replaced by the clinical records which make the things complex and difficult to take the advantage of the data in the database. No validation of the result or the final output is being done in the above presented algorithm.

Problem Statement

1. In our work we extend the work *"Coronary heart disease diagnosis by artificial neural networks including genetic polymorphisms and clinical parameters" by Oleg Yu. Atkov et al. (Elsevier, 2012)*.
2. Hospitals today adopted some hospital information system for handling as well managing their healthcare or patient as well clinical data.
3. Decision support systems already used by some hospitals already used, but are small in amount.
4. It can answer queries which are in simple forms like "What is the average age in years of patients who have heart disease with high BP rate?" and "How many of heart disease surgeries had resulted in hospital stays longer than 35 days?","Number of percentage of the female patients who are single, having higher heart rate below 26 years old, and who have been treated for cancer."

Problem Statement Contn..

5. But, they cannot answer complex type of queries like “Clarify the required preoperative predictors that increase the length of hospital stay”, “Specify patient data on cancer, should treatment include radiation only, chemotherapy alone or both radiation and chemotherapy?”, “Given patient records, predict the probability of patients that having higher blood pressure and having high heart rate getting a heart disease.”

6. Decisions in the hospital are generally done with respect to experience and guidance of the doctors rather than on the richer, non-trivial, useful, mined knowledge that hidden in the database records.

Methodology

- The term Classification And Regression Tree (CART) examination is an umbrella term used to allude to both of the above techniques, initially presented by Breimanet al.

- Trees utilized for relapse and trees utilized for arrangement have a few likenesses - additionally a few contrasts, for example, the system used to figure out where to part. It utilizes Gini polluting influences and data increase to figure choice tree.

- Another model that is used in this work is Artificial Neural Network, which is just similar to the human brain, the ANN is the Artificial Intelligence system which actually works on the basis of the past data

Methodology Contn…

- Considering an example to understand the ANN as the eyes captures the image of any object and send it to the brain network through neurons and then the neural network just matches the images captured by eyes with the already available images in the database of the human brain.

- If the images are perfectly matching then answer is that the object in front is an apple else like might be an apple or don’t know what it is if no such image found in the brain. Know in the case of medical database the ANN will be used for the pattern matching and also validating the results of diagnosis.

Methodology Contn…

- For example diagnosis of the patient with symptoms of cancer is to be done for that some short of sample and other related details are taken from the patient .

- On the basis of the range of the results of the sample and patient record matching of the patient pattern and pattern already available in the database is done on the basis of which the output is provided whether the patient is suffering from cancer or not.

- Also validating the same by showing the past records as the sample and record of the patient is actually matching the pattern in the database and for that we say whether the patient is having the diseases or not.
- http://www.sciencedirect.com/science/article/pii/S0914508711002255

Literature References

http://www.sciencedirect.com/science/article/pii/S0914508711002255

https://en.wikibooks.org/wiki/Artificial_Neural_Networks

https://www.extremetech.com/extreme/215170-artificial-neural-networks-are-changing-the-world-what-are-they

http://www.datasciencecentral.com/profiles/blogs/introduction-to-classification-regression-trees-cart

Step by Step processing of the proposed work

Step 1: Dataset Collection may be the fixed dataset from any organization or available online.

Step 2:Processing of the available dataset using the CART (Classification and Regression Tree) for data mining.

Step 3: Data collection and processing of data from patient by the means of form filling or soft methodology and processing of the same is being done using the CART.

Step 4:Pattern matching is being done using the ANN which actually works on the past data. Patterns here represents the records of the patient and records already existing in the database, patterns are been obtained by processing the data using the CART algorithm.

Step by Step processing of the proposed work Contn..

Step 5: Decision making processing using the ANN (Artificial Neural Network) methodology.

Step 6: Validation of the decision made on the basis of the past data is done using the Artificial Neural Network.

Step 7: Final result to the patient for the diseases for which he/she is to be treated.

Working Schedule of the Research Work

The complete dissertation is divided into following parts as Problem Generation Research methodology, Literature Review, Implementation, Analysis and Conclusion.

Month 1: (Jul'-Aug' 2017)

✓Reading Domain Related research papers,
✓Problem Identification/Formulation,
✓Synopsis writing.

Month 2: (Sept' 2017)

✓Reviewing different related methodologies,
✓Chapter 2: Literature Review write up,
✓Initialization of write up of chapter 1 Introduction.

Working Schedule Contn..

Month 3: (Oct' 2017)
Write up of complete chapter 1: Introduction
Working on research methodology after going related papers,
Chapter 3: Research Methodology Write up.

Month 4: (Implementation)
Coding of the Research methodology.

Month 5: (Analysis or Result Discussion)
Coding of Research Methodology in Python,
Analysis of the results with the already existing methodology,
Chapter 4: Result and Discussion write up,
Final bind up of the complete report.

Working Schedule Contn..

Month 3: (Oct' 2017)

✓Write up of complete chapter 1: Introduction
✓Working on research methodology after going related papers,
✓Chapter 3: Research Methodology Write up.

Month 4: (Implementation)

✓Coding of the Research methodology.

Month 5: (Analysis or Result Discussion)

✓Coding of Research Methodology in Python,
✓Analysis of the results with the already existing methodology,
✓Chapter 4: Result and Discussion write up,
✓Final bind up of the complete report.


THANK
YOU