



A Survey of Machine Learning Approaches and Health Diagnose Systems

Dr. Amjan Shaik

Professor of Information Technology,
BVRIT Narsapur, Medak, Telangana State, India

Abstract: *The Healthcare industry collects huge amounts of healthcare data which, unfortunately, are not “mined” to discover hidden information for effective decision making. Discovery of hidden patterns and relationships often goes unexploited. Advanced data mining techniques can help remedy this situation*

The Healthcare industry collects huge amounts of healthcare data which, unfortunately, are not “mined” to discover hidden information for effective decision making. Discovery of hidden patterns and relationships often goes unexploited. Advanced data mining techniques can help remedy this situation.

Data mining is often used during the knowledge discovery process and is one of the most important subfields in knowledge management. Data mining aims to analyze a set of given data or information in order to identify novel and potentially useful patterns. These techniques, such as Bayesian models, decision trees, artificial neural networks, associate rule mining, and genetic algorithms, are often used to discover patterns or knowledge that are previously unknown to the system and the users.

Data mining has been used in many applications such as health care, marketing, customer relationship management, engineering, medicine, crime analysis, expert prediction, Web mining, and mobile computing, among others. The development of Information Technology has generated large amount of databases and huge data in various areas. The research in databases and information technology has given rise to an approach to store and manipulate this precious data for further decision making.

Data mining is a process of extraction of useful information and patterns from huge data. It is also called as knowledge discovery process, knowledge mining from data, knowledge extraction or data /pattern analysis. Data mining is a logical process that is used to search through large amount of data in order to find useful data. The goal of this technique is to find patterns that were previously unknown. Once these patterns are found they can further be used to make certain decisions for development of their businesses.

Three steps involved are

1. Exploration
2. Pattern identification
3. Deployment

In data exploration, data is cleaned and transformed into another form, and important variables and then nature of data based on the problem are determined. Once data is explored, refined and defined for the specific variables the second step is to form pattern identification. Identify and choose the patterns which make the best prediction. Patterns are deployed for desired outcome.

Machine learning algorithms can be classified as supervised learning or unsupervised learning. In supervised learning, training examples consist of input/output pair patterns. Learning algorithms aim to predict output values of new examples based on their input values. In unsupervised learning, training examples contain only the input patterns and no explicit target output is associated with each input. The unsupervised learning algorithms need to use the input values to discover meaningful associations or patterns. Many successful machine learning systems have been developed over the past three decades in the computer science and statistics communities. The categorized five major paradigms of machine learning research, namely probabilistic and statistical models, symbolic learning and rule induction, neural networks, evolution-based models, and analytic learning and fuzzy logic.

The historical progress of machine learning and its applications in medical diagnosis shows that from simple and straight forward to use algorithms, systems and methodology have emerged that enable advanced and sophisticated data analysis. Data mining has importance regarding finding the patterns, forecasting, discovery of knowledge in different domains. Data mining techniques and algorithms such as classification, clustering helps in finding the patterns to decide upon the future businesses trends to grow. Data mining has wide application domain almost in every industry where the data is generated that's why data mining is considered one of the most important frontiers in database and information systems and one of the most promising interdisciplinary developments in Information Technology. In this review a broad overview of some of the data mining techniques, their use in various emerging algorithms and applications and also the protagonist of these in Health and Medical Systems.