A Bayesian Network Model for Probability Estimation

D

Harleen Kaur
Hamdard University, India

Ritu Chauhan
Amity University, India

Siri Krishan Wasan
Jamia Millia Islamia, India

INTRODUCTION

Deriving information from mountain of heathcare data can be too complex and voluminous to be processed by human capabilities alone. To overcome this flaw or nature of data, healthcare practioners are adopting new emergent technology to discover effective and efficient pattterns. However, medical researchers are exploiting numerous data mining techniques to find correlations or patterns among large scale databases as compared to traditional techniques for future medical diagnosis.

Data mining can be termed as an essential component for knowledge discovery process, where the vivid process determines effective and potential benefits among the data. However challenge is to establish proper exploitation strategy among health care data records which offer numerous facts in creation, dissemination and preservation of knowledge using advanced technologies, whereas, if the discovered knowledge tends to be a successful activity then gradually it can be used for futuristic decision making in healthcare organization. For instance if data of cancer patients or other diseases might consists of knowledgeable patterns which can be more expected to develop a kind of disease, so such knowledge can be used to prevent the diagnosis of patient's disease for futuristic decision making.

Progressively Data mining and knowledge discovery are used as interdisciplinary terms to discover hidden and unknown information from large scale databases. Eventually data mining is also known as essential step in knowledge discovery process, knowledge discovery process includes several integrated preprocessing and post processing steps to discover hidden information from databases. There exists several application domain areas of data mining techniques such as medical domain for diagnosis, management survey of data, marketing area of research, statistical analysis, and geographical analysis of data and several other research areas (Arabie & Hubert, 1994; Dunham, 2003; Kaur et al., 2010; Chauhan & Kaur, 2014). Data mining techniques are highly computational techniques under certain computational circumstances to retrieve effective and efficient patterns from raw data (Fayyad et al., 1996). The output of data mining techniques can be further applied for decision making support system, to retrieve profitable environment for experts and finally provide benefits to end users. Such analyses are increasing pressure on healthcare organizations to make decision based on data mining techniques to gain insights of data. Data mining can influence medical decision making by maintaining high level of healthcare.

Medical decision making from diagnosis to patient management is becoming more and more complex due to rapid growth of knowledge during last three decades. It is possible that even with specialization and super specialization physician may not be able to make an optimal decision. Computer assisted Medical Decision making use of data mining techniques may provide a partial solution to the problem. Since medical diagnosis is probabilistic in nature, it is well suited for probabilistic formalism. Bayesian classifiers are statistical classifiers based on famous Bayes theorem of conditional probability. Thus medical diagnosis fits well into Bayesian probabilistic framework. But there

DOI: 10.4018/978-1-4666-5888-2.ch148

are certain inherent limitations in Bayesian classification. For example, in medical applications this assumes that patient's symptoms are independent and patient has only one disease. Moreover, exact probabilities for various attributes responsible for a disease are required. Most of the above assumptions are not true in real life, for example, a patient may have several diseases at a time. The major challenge for the applications of Bayesian networks for medical diagnosis is to represent the domain knowledge in a probabilistic formalism.

We not only need medical data of patients but we must have ability to reason. What should be there in the diagnosis? This article examines Bayesian classification technique for exploration of medical diagnosis for various attributes resulting in development of futuristic decision making. The proper exploitation of Bayesian strategy can be deployed for further investigation and refinement of medical diagnosis. In this article we have utilized Bayesian classification technique for patients admitted to hospital for Right Heart Catheterization and used probability technique to evaluate different characteristics or features with sustainability rate among the patients.

BACKGROUND

Bayesian classifiers are also known as naïve Bayesian classifiers which are comparable in performance with decision tree and artificial neural networks (ANN). Bayesian belief networks are graphical models, which can also be used for classification. However Computer based medical diagnosis based on Bayesian techniques are developed for diagnosis of congenital heart disease by (Warner et al., 1961). The Bayesian techniques are also utilized for other medical diagnosis such as classification of strokes, ECG stress testing and coronary heart disease.

(Cooper & Herskovits, 1992) presented a Bayesian method for constructing a probabilistic network from a database where numerous cases were evaluated and demonstrated which can provide insight into probabilistic dependencies existing among the case variables. There exists no uncertainty among the probabilistic nature of cardiovascular disease; fits well into Bayesian-probabilistic framework but the complexities of this disease require other kind of reasoning.

(Long et al., 1994) have considered challenges collecting medical data and its presentation to the physician for appropriate diagnosis. Cardiovascular disease, provide a wide range of characteristic and disorders range from acute to chronic. The disease can progress and complicated other additional diseases for prolonged illness. (Long et al., 1994) used modified Bayesian Probability Network (BPN) to reason explanation among data and model casual path physiology for cardiovascular disease. Developing a BPN for diagnosis has many limitations. Bayesian classification assumes that patient's syndrome tends to be independent among each other and hence require probabilities in relation to all the attributes responsible for a particular disease. Obtaining all relative information about a particular patient, at times, may not be possible. (Long et al., 1994) have explained the problem for modeling heart disease types such as primary aortic regurgitation (AR) which can have different etiologies and may require reasoning for time domain. For example, acute myocardial Infarction (MI) could not explain pleural effusion on the same day as pulmonary congestion occasionally causes pleural effusion. One can learn a lot in terms of disease patterns; therefore it is important to capture experience of experts.

(Cooper and Herskovitis, 1991) described a Bayesian approach to learning the qualitative and quantitative dependency relationship among a set of discrete variables and called it as Bayesian Learning of belief network (BLN). Numerous medical approach has been utilized to develop methods for automated learning among data in field of statistics and artificial intelligence (Blum, 1982; Carbonell, 1989; Dagher & Herskovitis, 1996; Glymour, 1986; Hinton, 1990; Lele, 1988; James, 1985; Michalski et al., 1983). One can bridge BLN to other AI methods to form a basis for application to ANN and other data mining techniques to evolve medical diagnosis for heart diseases, in particular and other disease in general.

With the advent of sophisticated electronic data repositories, enormous amount of data in medical domain can be stored and useful knowledge can be retrieved utilizing several data mining methods. We may identify different characteristics of data which are accountable for a particular disease using numerous statistical methods. However appropriate Bayesian methods provide insight into probabilistic dependencies