Chapter 40
Comparative Study of Classification Models with Genetic Search Based Feature Selection Technique

Sanat Kumar Sahu
Department of Computer Science, Govt. K.P.G. College Jagdalpur, India

A. K. Shrivas
Department of IT, Dr. C. V. Raman University, Bilaspur, India

ABSTRACT

Feature selection plays a very important role to retrieve the relevant features from datasets and computationally improves the performance of a model. The objective of this study is to evaluate the most important features of a chronic kidney disease (CKD) dataset and diagnose the CKD problem. In this research work, the authors have used a genetic search with the Wrapper Subset Evaluator method for feature selection to increase the overall performance of the classification model. They have also used Bayes Network, Classification and Regression Tree (CART), Radial Basis Function Network (RBFN) and J48 classifier for classification of CKD and non-CKD data. The proposed genetic search based feature selection technique (GSBFST) selects the best features from CKD dataset and compares the performance of classifiers with proposed and existing genetic search feature selection techniques (FSTs). All classification models give the better result with proposed GSBFST as compared to without FST and existing genetic search FSTs.

INTRODUCTION

CKD is a serious health problem, affecting approximately 10% of the population worldwide (Subas, Alickovic, & Kevric, 2017). CKD or chronic renal disease progresses slowly and, generally, after months or years, the kidney loses its functionality (Polat, Danaei Mehr, & Cetin, 2017). In common, it may not

DOI: 10.4018/978-1-7998-2460-2.ch040

be detected before losing 25% of its functionality. In the Initial stage, CKD problems may not be predicted easily because may not give any initial symptoms. The undetected CKD is the risk of developing a progressive loss of renal function that can lead to renal failure, which means that regular dialysis or kidney transplantation is required to survive ("Chronic Kidney disease," n.d.). Yet, there's very little evidence on however CKD is diagnosed in a very efficient and automatic manner. Machine learning is a subdivision of artificial intelligence. This is the science of working without computer programmed explicitly (Haykin, 2008). It is related to the creation and study of systems that can learn from statistics. This paper examines how CKD can be diagnosed with the help of machine learning (ML) techniques. ML techniques have been powerful tools in the detection of anomalies in various physiological data. Various types of health dataset are classified with the help of ML in easily, and ML is one of the most successful tools. The most important purpose of solving classification problems, based on the quantity of attributes connected to the respective object, to determine the solution of problems that determine different types of measures to determine the objects in the predefined classes (Han, Kamber, & Pei, 2012; Pujari, 2013). Finding solutions to problems is whether or not the object belongs to a specific class. A number of attributes (feature) selection technique (Yildirim, 2015) include the machine learning area.The key role of these procedures is to eliminate irrelevant or redundant features from the dataset. Extraction in the reduced set of features has increased benefits (Jantawan & Tsai, 2014). It reduces the number of functions that appear in the models discovered, to help and make the models easier to understand. In this classification work, four different classification algorithms have been used. That is the CART, RBFN, J48, and BayesNet which used to categorize the CKD dataset. One Genetic Search (GS) (Hall, 1999) or Simple Genetic Algorithm (GA) is evolutionary search strategies based on the process of natural selection was used for FST. A comparative study of one feature selection techniques Genetic Search (GS) and proposed Genetic Search Based Feature Selection Techniques (GSBFST) were performed on the basis of the performance of the classification algorithms J48, RBFN, CART and Bayes Network for predicting the risks of CKD.

LITERATURE SURVEY

This part consists of reviews of various technical and related articles on machine learning techniques applied to predict kidney disease.

The two types (Polat et al., 2017) of feature selection methods, i.e., wrapper and filter approach have been used to diagnose CKD. In wrapper approach, a classifier subset evaluator with the greedy stepwise search engine and wrapper subset evaluator with the Best First Search(BFS) engine were used. In filter approach, correlation feature selection subset evaluator with a greedy stepwise search engine and filtered subset evaluator with the BFS engine were used. Results showed that the Support Vector Machine (SVM) classifier has used filtered subset evaluator with the BFS engine feature selection method gives a higher accuracy rate (98.5%) in the diagnosis of CKD.

A number of different ML classifiers (Subas et al., 2017) Artificial Neural Network(ANN), SVM, k-Nearest Neighbor, C4.5 and Random Forest (RF) have experiment validated to a real data set, taken from the UCI Machine Learning Repository. The result reveals that the random forest (RF) classifier reaches the maximum performances on the classification of CKD.

ML techniques (Charleonnan et al., 2016) were introduced to estimate CKD. Four ML methods were used, including near-neighbors (KNN), SVM, logistic regression (LR), and decision tree. These models