Chapter 1
Developing an Effective Classification Model for Medical Data Analysis

Naeem Ahmed Mahoto
Mehran University of Engineering and Technology, Pakistan

Abdul Hafeez Babar
Mehran University of Engineering and Technology, Pakistan

ABSTRACT

The sparse nature of medical data makes knowledge discovery and prediction a complex task for analysis. Machine learning algorithms have produced promising results for diversified data. This chapter constructs the effective classification model for medical data analysis. In particular, nine classification models, namely Naïve Bayes, decision tree (i.e., J48 and Random Forest), multilayer perceptron, radial bias function, k-nearest neighbors, single conjunctive rule learner, support vector machine, and simple logistics have been applied for developing an effective model. Besides, classification models have also been used in conjunction with ensemble learning methods, since ensemble methods significantly increase the predictive outcomes of the classification models. The evaluation of classification models has been measured using accuracy, f-measure, precision, and recall metrics. The empirical results revealed that the combination of ensemble learning methods with classification models produces better predictions in comparison with sole classification model for the medical data.

INTRODUCTION

Health is the key element of human life. Healthcare industry is one of the largest and fastest growing industries and is the backbone of our society (Lakshmi, Haritha & Srkit, 2016). The healthcare agencies, equipped with modern technological healthcare services, store information regarding health-related issues (Shah, Meghji & Mahoto, 2018; Mahoto, Shaikh & Ansari, 2014). Thus, everyday massive amount of data is being stored. This data is a great source for finding knowledge discoveries in order to determine disease trends, risk factors, and prediction of severe cases, medical pathways and general health inter-

ventions (Mahoto, Shaikh & Chowdhry, 2015; Mahoto, Shaikh & Khuhawar, 2014). Data mining allows discovery of hidden patterns from collected database (Banu & Gomathy, 2014). A main challenge, being faced by healthcare facilities, is delivery service with quality (Palaniappan & Awang, 2008). The quality service provides better and correct treatments to patients and helps in effective recovery processes. Poor diagnostic and treatment decisions may lead to disastrous and thus, are rarely distracted. Data mining techniques have been largely applied in healthcare for several purposes in the existing scientific literature. For instance, medical pathways have been extracted in Antonelli et al. (2012A) and Antonelli et al. (2012B), clustering in Antonelli et al. (2013), classification in (Babar & Mahoto, 2018; Shaikh et al., 2015) and association rule analysis has been carried out in Antonelli et al. (2015).

The World Health Organization (WHO) estimation inspired this study to be carried out, since WHO estimated, till 2030, almost 23.6 million individuals will expire due to severe complications of heart related issues (W.H.O., 2018). The possible reasons behind such a large number may be because people are overloaded with work burden in numerous developing countries. Eventually, many people suffer from mental stress and several other health complications. Therefore, it is an important task to diagnose the health related issues proactively, accurately and efficiently (Jarad, Katkar, Shaikh, & Salve, 2015). Generally, diagnosis is carried out based on medical expert's experience and knowledge. However, discovering the heart related problems based on merely symptoms, physical examination and signs of patient body is the difficult assignment in medical domain. This, being a multifaceted problem, may lead towards wrong hypothesis and unforeseeable effects (Turabieh, 2016). Nowadays, healthcare facilities have provided numerous amounts of medical data about patients, their diseases, diagnosis, and hospital resources, & many other important attributes. The advancement of machine learning field has opened ways for handling such a massive and high-dimensional data.

The classification of medical data helps in predicting and getting insight information of treatment procedures as well as historical outcomes. Several attempts have been made for the development of classification models to address the predictive outcomes of the medical data. However, medical data is inherently complex due to its sparse nature. This leads towards open challenge of predicting medical data. The machine learning algorithms have dealt with medical data and also ensemble-learning methods have been reported in the existing literature. Unfortunately, due to diversified nature of medical data, each algorithm behaves differently with different dataset. Therefore, it would be very likely that an algorithm that outperforms at certain medical data; it may not produce promising results for the other medical dataset. However, the investigation of several classification models for a certain medical data may provide broader picture of each considered classification model, which has been taken into consideration in this chapter.

This chapter aims at developing an efficient classification model, which may help in medical data analysis. Particularly, machine-learning algorithms such as Naïve Bayes (NB), Radial Bias Function (RBF), Decision Tree (DT), Random Forest (RF), Single Conjunctive Rule Learner (SCRL), K-Nearest Neighbor (KNN) and Support Vector Machine (SVM) would be considered for experimental activities. The performance of considered classification algorithms has been measured by exploiting well-established matrices: accuracy, precision, Recall and F-measure.

The rest of this chapter is organized as follows: Section 2 briefly presents related work done in the area of machine learning techniques applied for medical data; Section 3 describes classification models being applied in this chapter. The performance measures being used for evaluation of classification models are reported in section 4. The experimental results and conclusion are described in section 5 and Section 6 respectively.

15 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/developing-an-effective-classification-model-for-medical-data-analysis/222137

Related Content

Descriptive Data Analytics on Dinesafe Data for Food Assessment and Evaluation Using R Programming Language: A Case Study on Toronto's Dinesafe Inspection and Disclosure System
Ajinkya Kunjir, Jugal Shahand Vikas Trikha (2021). *Handbook of Research on Engineering, Business, and Healthcare Applications of Data Science and Analytics (pp. 485-507).*
www.irma-international.org/chapter/descriptive-data-analytics-on-dinesafe-data-for-food-assessment-and-evaluation-using-r-programming-language/264325

Integrating Blockchain Platforms With Big Data Solutions for Regional Innovation Development
Leyla Ayvarovna Gamidullaeva, Vardan Mkrttchianand Alexey Finogeev (2022). *Research Anthology on Big Data Analytics, Architectures, and Applications (pp. 381-390).*
www.irma-international.org/chapter/integrating-blockchain-platforms-with-big-data-solutions-for-regional-innovation-development/290992

EMG-Based Mobile Assessment System for Neck and Shoulder Fatigue
Pei Lun Lai, Hsiu-Sen Chiangand Qi-An Huang (2017). *International Journal of Big Data and Analytics in Healthcare (pp. 39-50).*
www.irma-international.org/article/emg-based-mobile-assessment-system-for-neck-and-shoulder-fatigue/204447

An Experimental Data of Lithium-Ion Battery Time Series Analysis: ARIMA and SECTRAL Analysis
Liming Xie (2021). *International Journal of Data Analytics (pp. 1-26).*
www.irma-international.org/article/an-experimental-data-of-lithium-ion-battery-time-series-analysis/285465

Big Data Mining: A Forecast to the Future
V. Sucharita, P. Venkateswara Rao, A. Satya Kalyanand P. Rajarajeswari (2018). *Exploring the Convergence of Big Data and the Internet of Things (pp. 36-42).*
www.irma-international.org/chapter/big-data-mining/187890