AI & Soc
DOI 10.1007/s00146-013-0456-0

OPEN FORUM

Comparative analysis of machine learning techniques in prognosis of type II diabetes

Abid Sarwar · Vinod Sharma

Received: 11 May 2012 / Accepted: 20 March 2013

Abstract Artificial Intelligence (AI) is now a days gaining immense importance and is becoming a key technology in many fields ranging from banking industry, to travel industry, to communication industry, and to robotic industry. The use of AI in medical diagnosis too is becoming increasingly popular and has been widely used in the diagnosis of tumors, cancers, hepatitis, lung diseases, etc. Numerous algorithms have been designed that help in the process of decision making by analyzing the hidden patterns in previously held information. The main objective of this manuscript is to apply multiple algorithms to a problem in the domain of medical diagnosis and analyze their efficiency in predicting the results. The problem selected for the study is the diagnosis of diabetes. Authors have identified ten parameters that play an important role in diabetes and prepared a rich database of training data which served as the backbone of the prediction algorithms. Keeping in view this training data, authors implemented three algorithms [Naïve Bayes, artificial neural networks (ANN), and *K*-nearest neighbors (KNN)] and developed prediction models. To calculate the efficiency, the results of prediction system were compared with the actual medical diagnosis of the subjects. The results indicate that the ANN is the best predictor with the accuracy of about 96 % which was followed by Naïve Bayes networks having an accuracy of about 95 % and the KNN came to be the worst predictor having an accuracy of about 91 %.

A. Sarwar (✉) · V. Sharma
Department of Computer Science and IT, University of Jammu, Jammu, India
e-mail: sarwar.aabid@gmail.com

V. Sharma
e-mail: vinodsharma@jammuuniversity.in

Keywords Artificial Intelligence · Diabetes · Naive Bayes classifier · Artificial neural networks · *K*-nearest neighbor · Medical diagnosis

1 Introduction

Artificial Intelligence, which is sometimes also referred to as "Synthetic Intelligence" (Lindley 2012), is defined as the branch of science and engineering concerned with the computational understanding of what is commonly called intelligent behavior and with the creation of artifacts that exhibit such behavior. Programs which enable computers to function in the ways that make people seem intelligent are called artificial intelligent systems (Ramesh et al. 2004). The field was founded on the claim that a central property of humans, intelligence, can be simulated by a machine. Artificial Intelligence has now days become an essential part of the technology industry, providing the heavy lifting for many of the most difficult problems in computer science. The general aim of AI is simulating Human Intelligence, which can be broken down into a number of specific sub-specific aims/domains depending upon the particular traits or capabilities that researchers would like an intelligent system to display. These domains include machine learning, natural language processing, knowledge representation, etc. Machine learning which is one branch of AI aims at providing computational methods for accumulating, changing, and updating knowledge in the intelligent systems. Machine learning methods are useful in cases where algorithmic solutions are not available and there is a lack of formal models or the knowledge about the application domain is poorly developed. Machine learning algorithms were from the very beginning designed and used to analyze different data sets. A major focus of machine learning is the

Published online: 16 April 2013