Chapter 20
Using Ensemble Learning and Random Forest Techniques to Solve Complex Problems

V. Belsini GladShiya
Agurchand Manmull Jain College, India

K. K. Sharmila
VISTAS, India

ABSTRACT

The branch of computer science and artificial intelligence known as machine learning is used to program machines to learn. Algorithms for machine learning are software programs or methods used to find hidden patterns in data, predict outcomes, and improve performance based on past performance. A technique used in machine learning called ensemble learning combines several models, such as classifiers or experts that have been carefully constructed to solve a particular computational intelligence problem. Ensemble refers to a collaborative effort to create a single impact. An ensemble can predict events more accurately and perform better in general than a single contributor. A random forest is a technique for ensemble learning in which many decision trees are combined to create the forest. This chapter covers the fundamentals of ensemble learning using random forest, implementation with real-world examples, and developing a model.

INTRODUCTION

Machine Learning is the subfield of computer science and Artificial Intelligence to make the machines to learn. Artificial intelligence, which is widely defined as a machine's ability to mimic intelligent human behavior used to carry out complicated tasks in a manner how people solve issues. It is a modern inventions in business and professional procedures as well as our daily lives. It's a branch of artificial intelligence (AI) that focuses on creating intelligent computer systems that can learn from databases by employing statistical approaches. With the support of machine learning the computers automatically

DOI: 10.4018/978-1-6684-8531-6.ch020

Copyright © 2024, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

learn from data and previous experiences while seeing patterns to generate predictions with a minimum of human involvement. Computers can now function independently without explicit programming using machine learning techniques. ML apps may freely learn from fresh data and grow, develop, and adapt. Machine learning uses algorithms to find patterns and learn in an iterative process, extracting valuable knowledge from massive amounts of data. Machine Learning algorithms are programs or techniques used to discover hidden patterns in data, forecast results, and enhance performance based on past performance (Charbuty, B.,& Abdulazeez, A.M.,2021). In machine learning, several algorithms can be employed for various tasks, such as basic linear regression for prediction issues or categorization issues. The machine learning algorithms must draw a conclusion from observed values and decide which category new observations fall into while doing classification jobs (Wei Jin,2020). The machine learning algorithm must estimate and comprehend the relationships between the variables in regression problems. Regression analysis is very helpful for prediction and forecasting since it concentrates on one dependent variable and a number of other changing factors. A typical method for analyzing trends, forecasting entails generating predictions about the future based on facts from the past and present. The machine learning algorithms should be selected according to the need of the applications (Amir et al., 2019). The best machine learning algorithm depends on a number of variables, such as the quantity, quality, and diversity of the data, as well as the conclusions that organizations hope to draw from it. Accuracy, training duration, parameters, data points, and many other factors are also important. As a result, selecting the appropriate algorithm requires consideration of the business need, the specification, the experimentation, and the time available. Even the most seasoned data scientists are unable to predict which algorithm will perform the best without first testing alternatives. Hence number of machine learning algorithms is used for various applications (In Lee a, Yong Jae Shin,2019). Ensemble learning is a process in machine learning in which the number of models such as classifiers or experts, are strategically developed are merged to address a specific computational intelligence problem. Ensemble means producing a single effect with the group. The main goals of ensemble learning are to enhance categorization, prediction, and function approximation. The ensemble methods can be done by mapping operations that each contributing member has learned are combined in ensemble learning methods (Pintelas et al., 2020). The combination of decision boundaries of members provides the clearest understanding of ensembles for classification. The combination of member hyper planes provides the best understanding of ensembles for regression (Fatai Anifowose, 2020). The ensemble learning is used for two purposes such as Performance and Robustness. Compared to a single contributing model, an ensemble can anticipate events more accurately and perform better overall. An ensemble narrows the prediction and model performance distribution. The ensemble model also used to overcome noise, bias and variance to improve the overall performance. It also gives the solution to overcome the challenges occur in a single estimator.

There are various numbers of algorithms in machine learning and the algorithms are used to ensemble learning process. In this chapter the random forest algorithm is concerned and its methods of processing could be progressed. Random forest algorithm is used to solve both classification and regression problems. A “forest” is created by growing and combining various decision trees using the supervised machine learning method Random Forest(Onesmus Mbaabu, 2020). Random forest is an ensemble learning technique where more decision trees are ensemble to get the forest. The ensemble techniques used in random forest are bagging. The “bagging” approach utilizes a Bootstrap Aggregation ensemble machine learning technique. An ensemble technique combines predictions from various machine learning algorithms to provide predictions that are more precise than those from a single model. To create sample datasets for each model, Bootstrap randomly selects rows and features from the dataset. These

18 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/using-ensemble-learning-and-random-forest-techniques-to-solve-complex-problems/335200

Related Content

Robotics and Artificial Intelligence
Estifanos Tilahun Mihret (2020). *International Journal of Artificial Intelligence and Machine Learning (pp. 57-78).*
www.irma-international.org/article/robotics-and-artificial-intelligence/257272

A Privacy Protection Method for Deepfake Hybrid Cloud Computing
L. Subhashini, S. Maheswariand S. Prabhu (2023). *Handbook of Research on Advanced Practical Approaches to Deepfake Detection and Applications (pp. 120-132).*
www.irma-international.org/chapter/a-privacy-protection-method-for-deepfake-hybrid-cloud-computing/316748

Towards Integrating Data Mining With Knowledge-Based System for Diagnosis of Human Eye Diseases: The Case of an African Hospital
Nilamadhab Mishraand Johny Melese Samuel (2021). *Handbook of Research on Disease Prediction Through Data Analytics and Machine Learning (pp. 470-485).*
www.irma-international.org/chapter/towards-integrating-data-mining-with-knowledge-based-system-for-diagnosis-of-human-eye-diseases/263334

Designing a Real-Time Dashboard for Pandemic Management: COVID-19 Using Qlik Sense
Rahul Rai (2021). *Machine Learning and Data Analytics for Predicting, Managing, and Monitoring Disease (pp. 190-203).*
www.irma-international.org/chapter/designing-a-real-time-dashboard-for-pandemic-management/286252

Assessing Hyper Parameter Optimization and Speedup for Convolutional Neural Networks
Sajid Nazir, Shushma Patelandand Dilip Patel (2020). *International Journal of Artificial Intelligence and Machine Learning (pp. 1-17).*
www.irma-international.org/article/assessing-hyper-parameter-optimization-and-speedup-for-convolutional-neural-networks/257269