Ensemble Methods and Their Applications

M. Govindarajan
Annamalai University, India

INTRODUCTION

Considering the variety of ensemble techniques and the large number of combination schemes proposed in literature, it is not surprising that a very large number of ensemble methods and algorithms are now available to the research community. To help the researchers and practitioners to get their bearings and to develop new methods and techniques, several taxonomies of ensemble methods have been proposed. Indeed, ensemble methods are characterized by two basic features: 1) the algorithms by which different base learners are combined; 2) the techniques by which different and diverse base learners are generated. This chapter basically distinguishes between non-generative ensemble methods that mainly rely on the former feature of ensemble methods, and generative ensembles that mainly focus on the latter. It is worth noting that the “combination” and the “generation” of base learners are somehow both present in all ensemble methods: the distinction between these two large classes depends on the predominance of the combination or of the generation component of the ensemble algorithm. More precisely, non-generative ensemble methods confine themselves to combine a set of possibly well-designed base classifiers: they do not actively generate new base learners but try to combine in a suitable way a set of existing base classifiers. On the contrary, generative ensemble methods generate sets of base learners acting on the base learning algorithm or on the structure of the data set to try to actively improve diversity and accuracy of the base learners. In this case the emphasis is placed on the way diverse base learners are constructed, while the combination technique does not represent the main issue of the ensemble algorithm. The main aim of this chapter is to explain the detailed characteristics of each ensemble methods and to provide an overview of the main application areas of ensemble methods. The rest of the chapter is organized as follows: the background section describes the related work. A brief description of ensemble methods reported in the literature, distinguishing between generative and non-generative methods, main application areas of ensemble methods is presented in section of main focus of the chapter. Finally, the chapter concludes with future research directions.

BACKGROUND

Patil et al., (2016) attempts to build robust data mining models to predict the defaulters using data obtained from one of finance company and various ensemble algorithms like bagging, boosting and stacking are implemented. Araque et al., (2017) seek to improve the performance of deep learning techniques integrating them with traditional surface approaches based on manually extracted features. Chubato Wondaferaw Yohannese et al. (2018) are to conduct large scale comprehensive experiments to study the effect of resolving those challenges in Software Fault Prediction in three stages in order to improve the practice and performance of Software Fault Prediction. Silas Nzuva et al., (2019) review the superiority exhibited by ensemble learning algorithms based on the past that has been carried out over

DOI: 10.4018/978-1-7998-9220-5.ch109

Copyright © 2023, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

the years. Gupta et al., (2019) proposed an approach for Facial Expression Recognition using Ensemble Learning Technique. In the proposed method, initially the features are extracted from static images using color histograms. This process is done for all images gathered in the training dataset. The ensemble technique is then applied on the featured dataset in order to categorize a given image into one of the six emotions, happy, sad, fear, angry, disgust, and surprise. Peterson et al., (2019) introduce the application of canonical correlation analysis fusion as a method for water-based spectral analysis to overcome the low signal-to-noise ratio of the data. Water-quality variables and spectral reflectance were used to create predictive models via machine learning regression models, including multiple linear regression, partial least-squares regression, Gaussian process regression, support vector machine regression, and extreme learning machine regression. The models were then combined using decision-level fusion. Thanh and Lang (2019) analyzes and evaluates the performance of using known ensemble techniques such as Bagging, AdaBoost, Stacking, Decorate, Random Forest and Voting to detect DoS attacks on UNSW-NB15 dataset, created by the Australian Cyber Security Center 2015. Singh et al., (2020) have developed a model to combine the results of each individual base learner using Bagging and Boosting ensemble methods. The results obtained using bagging and boosting ensemble techniques were compared to select the best model. Yu et al., (2020) propose an end-to-end deep ensemble learning based on the weight optimization (DELWO) model. It contributes to fusing the deep information derived from multiple models automatically from the data. Huang et al., (2020) proposes an ensemble-learning-approach-based solution of integrating a rich body of features derived from high resolution satellite images, street-view images, building footprints, points-of-interest (POIs) and social media check-ins for the urban land use mapping task. The proposed approach can statistically differentiate the importance of input feature variables and provides a good explanation for the relationships between land cover, socioeconomic activities and land use categories. Rajagopal et al., (2020) present an ensemble model using meta classification approach enabled by stacked generalization. Akhter et al., (2021) use five ensemble learning methods to ensemble the base-predictors' predictions to improve the fake news detection system's overall performance. Gaikwad (2021) proposed a novel ensemble classifier using rule combination method for intrusion detection system. Ensemble classifier is designed using three rule learners as base classifiers. The benefits and feasibility of the proposed ensemble classifier have demonstrated by means of KDD'98 datasets. The main novelty of the proposed approach is based on three rule learner combination using rule of combination method of ensemble and feature selector. These three base classifiers are separately trained and combined using average probabilities rule combination. Mehmet Akif Yamana et al., (2021) presented different algorithms employing skin color for face detection, histogram tor the feature extraction and ensemble methods for recognition part. Hence, the developed Bagging and Boosting has thoroughly presented their superiority over FERET database. Advanced classification capabilities of the ensemble classifiers are having better accuracy than the existing methods in the face recognition. Jafarzadeh et al., (2021) investigates the capability of different EL algorithms, generally known as bagging and boosting algorithms, including Adaptive Boosting (AdaBoost), Gradient Boosting Machine (GBM), XGBoost, LightGBM, and Random Forest (RF), for the classification of Remote Sensing (RS) data.

14 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/ensemble-methods-and-their-applications/317587

Related Content

Comparison of Brainwave Sensors and Mental State Classifiers
Hironori Hiraishi (2022). *International Journal of Artificial Intelligence and Machine Learning (pp. 1-13).*
www.irma-international.org/article/comparison-of-brainwave-sensors-and-mental-state-classifiers/310933

Data Science in the Database: Using SQL for Data Preparation
Antonio Badia (2023). *Encyclopedia of Data Science and Machine Learning (pp. 1187-1200).*
www.irma-international.org/chapter/data-science-in-the-database/317523

Development of a Charge Estimator for Piezoelectric Actuators: A Radial Basis Function Approach
Morteza Mohammadzaheri, Mohammadreza Emadi, Mojtaba Ghodsi, Issam M. Bahadur, Musaab Zarogand Ashraf Saleem (2020). *International Journal of Artificial Intelligence and Machine Learning (pp. 31-44).*
www.irma-international.org/article/development-of-a-charge-estimator-for-piezoelectric-actuators/249251

Multilayer Neural Network Technique for Parsing the Natural Language Sentences
Manu Pratap Singh, Sukrati Chaturvediand Deepak D. Shudhalwar (2019). *International Journal of Artificial Intelligence and Machine Learning (pp. 22-38).*
www.irma-international.org/article/multilayer-neural-network-technique-for-parsing-the-natural-language-sentences/238126

Unveiling the World of Bioinformatics
Khushboo Dhimanand Hardik Dhiman (2024). *Applying Machine Learning Techniques to Bioinformatics: Few-Shot and Zero-Shot Methods (pp. 181-200).*
www.irma-international.org/chapter/unveiling-the-world-of-bioinformatics/342724