Ensemble Methods

Enrique Chi Gongora, *1909040*, Victor Moo Quintal, *2009098*, Brandon Pacheco Chan, *2009102*, Raymundo Baas Cabañas, *1909008*, Rafael Marí Reyna, *2009083*, Jose Pat Ramírez, *2009104*, Denzel Chuc Chuc, *st1809052*

***Abstract*—Ensemble methods represent an important area in machine learning, enhancing performance by combining multiple model predictions. These methods, which construct a set of base learners to create a composite model, aim to surpass individual model accuracy and robustness, particularly through the reduction of overfitting and the enhancement of generalization to new data. Distinguished by their ability to convert weak learners into strong ones, ensemble methods have evolved through contributions from various domains such as pattern recognition and neural networks. Since the 1990s, these methods have been solidified as a significant learning paradigm, substantiated by both empirical evidence and theoretical proofs. Applications in fields ranging from computer vision to medical diagnosis highlight the versatility and effectiveness of ensemble methods. Techniques such as bagging, boosting, stacking, and voting showcase the adaptability of these methods to different data and problem structures, thus ensuring improved and reliable model performance. As ensemble methods continue to advance, they promise innovative solutions to complex and emerging problems, underscoring their integral role in the future of machine learning.**

I. Introduction

Ensemble methods are a fundamental pillar of machine learning strategies that improve overall performance by integrating decisions from multiple models. These methods construct a set of base learners from the training data and then build a composite model that makes the final prediction based on the predictions of these base learners. The primary objective of ensemble methods is to amalgamate the predictions of several base learning algorithms, thereby producing a model that surpasses the individual models in terms of accuracy and robustness. Designed to enhance the stability and predictive power of machine learning algorithms, ensemble methods have found successful applications across a diverse range of fields, including computer vision, natural language processing, and bioinformatics. The central premise behind ensemble methods is the concept that a more robust and generalizable model can be constructed by combining multiple weak models. This makes ensemble methods a potent tool for addressing complex machine learning problems, reducing the likelihood of overfitting and improving the model's ability to generalize from the training data to unseen data. By harnessing the power of multiple learning models, ensemble methods effectively capture a more comprehensive understanding of the data patterns, leading to superior predictive performance.

II. Development

A. Ensemble Methods

Ensemble methods are a type of machine learning approach that train multiple learners, or base learners, to solve the same problem. Unlike traditional learning approaches that construct a single learner from training data, ensemble methods construct a set of learners and combine them. [1]

Base learners can be generated from training data by a base learning algorithm, such as a decision tree, neural network, or other types of learning algorithms. Most ensemble methods use a single base learning algorithm to produce homogeneous base learners, i.e., learners of the same type, leading to homogeneous ensembles.

The generalization ability of an ensemble is often much stronger than that of base learners. In fact, ensemble methods are appealing mainly because they are able to boost weak learners, which are just slightly better than random guess, to strong learners which can make very accurate predictions.

There are three threads of early contributions that led to the current area of ensemble methods: combining classifiers, ensembles of weak learners, and mixture of experts. Combining classifiers was mostly studied in the pattern recognition community, ensembles of weak learners was mostly studied in the machine learning community, and mixture of experts was mostly studied in the neural networks community.

Ensemble methods have become a major learning paradigm since the 1990s, with great promotion by two pieces of pioneering work. One is empirical, in which it was found that predictions made by the combination of a set of classifiers are often more accurate than predictions made by the best single classifier. The other is theoretical, in which it was proved that weak learners can be boosted to strong learners.

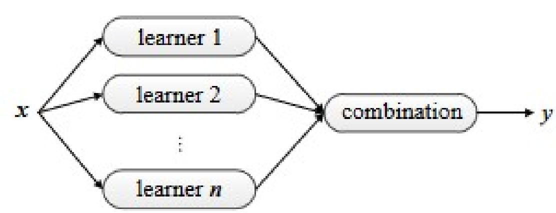

Fig. 1. Common Ensemble Architecture

B. Applications of Ensemble Methods

[1] Ensemble methods have been successfully applied in a variety of fields, demonstrating their versatility and effectiveness. In the realm of object tracking, an innovative approach known as ensemble tracking has been proposed. This method trains an ensemble online to distinguish between the object and the background, constantly updating a set of weak classifiers to incorporate new information about changes in object appearance and the background.

In computer security, ensemble methods have proven to be particularly effective. Each activity performed on computer systems can be observed at multiple abstraction levels, and relevant information may be collected from multiple information sources. Ensemble methods have been applied to intrusion detection, constructing an ensemble from each type of features independently, and then combining the outputs from these ensembles to produce the final decision.

The detection of malicious executables is another area where ensemble methods have shown promise. An ensemble method has been proposed to detect previously unseen malicious executables automatically, based on representing the programs using binary profiling, string sequences, and hex dumps.

In the medical field, ensemble methods have been instrumental in increasing the reliability of diagnoses. A two-layered ensemble architecture has been designed for lung cancer cell identification. For the early diagnosis of Alzheimer's disease, an ensemble method has been proposed where the component learners are trained on different data sources obtained from different electrodes in response to different stimuli and in different frequency bands, and their outputs are combined for the final diagnosis.

These applications underscore the power and flexibility of ensemble methods for addressing complex problems in a variety of domains. The adaptability of ensemble methods allows them to be applied in diverse fields, from object tracking in computer vision to intrusion detection in computer security, and from detecting malicious executables to aiding in medical diagnoses. This broad applicability is a testament to the robustness of ensemble methods. They can handle different types of data, adapt to various problem structures, and provide reliable and improved performance. The success of ensemble methods in these areas highlights their potential for future applications in emerging fields, promising a new wave of innovative solutions to complex problems. Their ability to boost weak learners and combine multiple models to make accurate predictions makes them a powerful tool in the arsenal of machine learning techniques. As more progress is made in this area, we can expect to see even more sophisticated applications of ensemble methods, further demonstrating their value in solving real-world problems.

C. Types Of Ensamble Methods

1) Bagging: Bagging, which stands for Bootstrap Aggregation, is an ensemble technique where a single training algorithm is used on different subsets of the training data. These subsets are created using a technique called bootstrapping, which involves sampling different sets of data from a given training set with replacement. The algorithm is trained on all the subsets, and the results are then aggregated. This aggregation typically involves majority voting for classification problems and averaging for regression problems.

The main advantage of bagging is that it significantly decreases the variance without increasing bias. This is largely due to the diversity in the training data, as the sampling is done by bootstrapping. Furthermore, if the training set is very large, bagging can save computational time by training the model on a smaller data set, while still increasing the accuracy of the model. It also works well with small datasets.

However, bagging does have some disadvantages. It improves the accuracy of the model at the expense of interpretability. For instance, if a single tree was being used as the base model, it would have a more easily interpretable diagram, but with the use of bagging, this interpretability gets lost. Another disadvantage is that during sampling, it's not clear which features are being selected. There's a chance that some features are never used, which could result in a loss of important information.

In bagging, there might be some data that are never sampled at all. These remaining data, which are not sampled, are called out of bag instances. The Random Forest approach is a notable example of a bagging method where deep trees, fitted on bootstrap samples, are combined to produce an output with lower variance. [2]

2) Boosting: It is an ensemble technique that converts weak learners into strong ones. It involves training a single algorithm on different subsets of the training data, with each subset focusing on the observations that were poorly classified by the previous learners. This sequential training and aggregation of results help to make accurate predictions on all kinds of data, not just the most common or easy observations.

In boosting, the dataset is weighted so that observations that were incorrectly classified by a classifier are given more importance in the training of the next model. While in bagging the weak learners are trained in parallel using randomness, in boosting the learners are trained sequentially, with each subsequent learner aiming to reduce the errors of the previous learners. Boosting can be used for regression as well as for classification problems and is mainly focused on reducing bias.

The choice of trees as a base for the boosting technique is due to their computational scalability, ability to handle missing values, robustness to outliers, ability to deal with irrelevant inputs, and interpretability. However, trees have the inability to extract a linear combination of features and high variance leading to a small computational power. Boosting helps to minimize the variance by taking into consideration the results from various trees.

Boosting has several advantages, including its resilience that curbs over-fitting easily, its proven effectiveness, and its versatility as it can be applied to a wide variety of problems. However, it also has some disadvantages. It is sensitive to outliers since every classifier is obliged to fix the errors in the predecessors, making the method too dependent on outliers. Another disadvantage is that the method is almost impossible to scale up because every estimator bases its correctness on the previous predictors, making the procedure difficult to streamline. Common examples of Boosting Techniques include Ada Boost (Adaptive Boosting), Gradient Boosting, and XG Boost (Xtreme Gradient Boosting). [3]

3) Stacking: Stacked Generalization, or "Stacking", is an ensemble machine learning algorithm that combines multiple machine learning algorithms via meta-learning. Unlike bagging and boosting which often consider homogeneous weak learners, stacking often considers heterogeneous weak learners, meaning different learning algorithms are combined. Additionally, stacking learns to combine the base models using a meta-model, whereas bagging and boosting combine weak learners following deterministic algorithms.

In stacking, base level algorithms are trained based on a complete training dataset, and the meta-model is trained on the final outcomes of all base-level models as a feature. Each classifier in stacking works independently, which permits the classifiers with different hypotheses and algorithms. For instance, a model can be trained on linear regression classifier, Decision Tree, and Random Forest for training, and then their predictions can be combined using a Support Vector Machine.

The benefit of stacking is that it can harness the capabilities of a range of well-performing models on a classification or regression task and make predictions that have better performance than any single model in the ensemble. Stacking, just like other ensemble techniques, tries to improve the accuracy of a model by using predictions of not so good models and then using those predictions as an input feature for a better model.

However, stacking does have a disadvantage. As the whole dataset is used for training for every individual classifier, in the case of huge datasets, the computational time will be more as each classifier is working independently on the huge dataset. Despite this, stacking improves the model prediction accuracy and is a powerful tool in the arsenal of machine learning techniques.

4) Voting: It is a fundamental combination method in ensemble learning, particularly useful when dealing with nominal outputs. It involves each classifier voting for one class label, and the final output class label is the one that receives the most votes. There are several types of voting methods, including majority voting, plurality voting, and weighted voting.

Majority voting requires the final winner to take at least half of the votes. If none of the class labels receives more than half of the votes, a rejection option is given and the combined classifier makes no prediction. Plurality voting, on the other hand, takes the class label which receives the largest number of votes as the final winner.

Weighted voting assigns a weight to each classifier, and the combined output for a class is the sum of the product of the weight and the output of each classifier for that class. The weights can be classifier-specific, class-specific, or assigned to each example of each class for each classifier.

Soft voting is generally used for classifiers which produce class probability outputs. In this method, the individual classifier outputs a vector for an instance, which can be regarded as an estimate of the posterior probability. The simple soft voting method generates the combined output by simply averaging all the individual outputs. If we consider combining the individual outputs with different weights, the weighted soft voting method can take several forms, similar to weighted voting.

These voting methods in ensemble learning not only improve the accuracy of the model but also enhance the robustness and reliability of the predictions.

D. Combination Methods

Combination methods in ensemble learning offer significant benefits. They address three fundamental issues that traditional learning approaches often fail to tackle: statistical, computational, and representational.

Statistical Issue: In many scenarios, the hypothesis space is too large to explore for limited training data, and there may be several different hypotheses giving the same accuracy on the training data. Combination methods reduce the risk of choosing a wrong hypothesis by combining the hypotheses.

Computational Issue: Many learning algorithms perform some kind of local search that may get stuck in local optima. By running the local search from many different starting points, the combination may provide a better approximation to the true unknown hypothesis.

Representational Issue: In many machine learning tasks, the true unknown hypothesis could not be represented by any hypothesis in the hypothesis space. By combining the hypotheses, it may be possible to expand the space of representable functions, and thus the learning algorithm may be able to form a more accurate approximation to the true unknown hypothesis.

Through combination, the variance as well as the bias of learning algorithms may be reduced. This has been confirmed by many empirical studies. Therefore, combination methods in ensemble learning not only improve the accuracy of the model but also enhance the robustness and reliability of the predictions.

E. Model Complexity and Ensemble Size in Ensemble Learning

The efficiency of ensemble methods is often measured by their predictive performance, which can be influenced by the complexity of base models and the size of the ensemble. Ensemble learning contends with the principle of Occam's Razor, which posits that simpler models with fewer assumptions are generally preferable over complex ones. In the context of ensemble learning, this principle raises important considerations regarding the balance between model complexity and predictive accuracy. Studies suggest that while ensembles composed of numerous complex models may offer increased accuracy, they also introduce a greater number of assumptions, potentially leading to overfitting and reduced model generalizability. This complexity must be managed to maintain the ensemble's ability to generalize beyond the training data [4].

F. Performance Benefits of Ensemble Learning Models

The efficacy of ensemble learning models is significantly enhanced when applied to complex nonlinear data. It has been observed that ensemble models exhibit superior performance over single learner models, particularly in scenarios involving nonlinear relationships within the data [5]. The ensemble learning approach capitalizes on its inherent ability to synthesize the predictive capabilities of multiple base learners, each contributing to a more nuanced understanding and representation of the complex patterns present in nonlinear datasets.

A study conducted on the application of ensemble learning models to geotechnical data further substantiates the notion that ensemble methods offer significant advantages. The study highlights that ensemble models, such as CatBoost and Random Forest, yield better performance in comparison to commonly used single learners. This finding is crucial as it confirms the premise that ensemble learning is not only beneficial but may be essential for accurately processing and predicting outcomes from complex and nonlinear data structures [5].

The advantages afforded by ensemble learning in such contexts are attributed to its robustness and the collective intelligence of multiple models. The ensemble's aggregated predictions are less susceptible to the idiosyncrasies of individual base learners, and thus, are more adept at capturing the true underlying patterns of the data [5].

III. Conclusion

Ensemble learning has proven to be a strong and effective approach in the field of machine learning. By bringing together the strengths of different models, it creates a combined predictive power that is often greater than any single model alone. This method works similarly to how a group of experts would come together to make a decision, with each model contributing its unique perspective to arrive at a more accurate and well-rounded conclusion. This collective approach helps to avoid common problems such as overfitting, especially when dealing with data that may not vary much.

Techniques like cross-validation are used to make sure each model plays a useful role in the final outcome. The versatility of ensemble learning is evident in its use of various methods like bagging, boosting, stacking, and voting, each helping to lower errors and improve the accuracy of predictions. These methods have been applied to a wide range of problems, providing reliable results even when the data changes or new challenges arise. Ensemble learning continues to be a key part of machine learning, pushing forward the capabilities of predictive modeling and classification.

References

[1] Z.-H. Zhou, Ensemble Methods: Foundations and Algorithms. Taylor Francis Group, 2012.

[2] I. Ahmad, M. Yousaf, S. Yousaf, and M. O. Ahmad, ”Fake News Detection Using Machine Learning Ensemble Methods,” in Complexity, vol. 2020, pp. 1-11, Oct. 2020. Accessed on: Nov. 8, 2023. [Online]. Available: https://doi.org/10.1155/2020/8885861

[3] A. Chauhan, “Ensemble methods - bagging, boosting, and stacking,” Medium, https://medium.com/analytics-vidhya/ensemble-methods-bagging-boosting-and-stacking-28d006708731 [accessed Nov. 6, 2023].

[4] J. Brownlee, ”Ensemble Learning Algorithm Complexity and Occam's Razor,” Machine Learning Mastery, April 27, 2021. [Online]. Available: https://machinelearningmastery.com/ensemble-learning-and-occams-razor/#:~:text=By%20Jason%20Brownlee%20on%20April,will%20make%20them%20too [Accessed: 9-November-2023].

[5] A. Li and J. Zhang, ”A comprehensive evaluation of ensemble machine learning in geotechnical engineering: Case study of a real-world earth-fill dam,” Geotechnical and Geological Engineering, vol. 41, no. 1, pp. 1-18, 2023. [Online]. Available: https://link.springer.com/article/10.1007/s10999-023-09679-0 [Accessed: 9-November-2023].