

Title: Understanding the Difference between Decision Tree and Random Forest algorithms

A decision tree classifier (DTC) is a well-known method for data classification. The most significant feature of DTC is its ability to change complicated decision-making problems into simple processes, thus finding a solution that is understandable and easier to interpret. This paper provides a brief review of various algorithms developed in literature for constructing and representing decision trees, splitting criteria for selecting best attribute and pruning methods. The readers will be able to understand why decision trees are more popular among all other methods of classification, what are their uses, limitations and applications in different diverse areas. They will also come to know about a decision tree induction algorithms, splitting criteria, pruning methods, concepts of ensemble methods, fuzzy decision trees, hybridisation of DTCs, etc. These enhancements are found very helpful in solving complex datasets with less computation in very short time period while achieving high accuracy. (**Priyanka and Dharmender Kumar**)

Application programming interface(API) benefits to effectively improve software development efficiency by reusing existing software frameworks or libraries.However,many constraints must be satisfied to correctly use APIs,such as call order,exception handling.Violation of these constraints will cause API misuse,which may result in software crashes,errors,or vulnerabilities.Although many API misuse detection techniques have been proposed,these techniques still face two challenges:1) the acquisition of API usage specification is difficult,and 2) the detection of many different types of API misuse at the same time is difficult.To address the above challenges,a decision tree algorithm-based API misuse detection method is proposed.First,the API usage source code is converted into an API usage graph,and the API usage specification is mined from the graph to effectively solve the first challenge.Second,an API usage decision tree is constructed based on the obtained API specification information,and the generalization (**Ji suan ji ke xue**)

The general principle of random forests is to aggregate a collection of random decision trees. The goal is, instead of seeking to optimize a predictor “at once” as for a CART tree, to pool a set of predictors (not necessarily optimal). Since individual trees are randomly perturbed, the forest benefits from a more extensive exploration of the space of all possible tree predictors,

which, in practice, results in better predictive performance. Focusing on random forests, this chapter begins by addressing the instability of a tree and subsequently introduces readers to two random forest variants: Bagging and Random Forest Random Inputs. The construction of random forests is illustrated on the spam dataset using the randomForest package. The clever prediction error estimate Out-Of-Bag Error is also presented. In turn, the chapter assesses the sensitivity of prediction performance to the two main parameters: the number of trees and the number of variables picked at each node. In the final section, random forests are applied to three examples: predicting ozone concentration, analyzing genomic data, and analyzing dust pollution. **(Random Forests book)**

❖ Advantage

Media diversity. The internet is a multipurpose means of conveying information to learners around the world. Internet sites may contain various media, including text, audio, graphics, animation, video, and downloadable software. New information. Now, with the ability to connect to resources in the community and around the world, students can access libraries and databases which are frequently updated on a daily basis. E-learning increases opportunities for smaller schools as well as for individuals to participate in home schooling. (1) Navigation. The main advantage of the internet is the ability to move easily within and between documents. With the press of a button or the click of a mouse, users can search for documents in multiple locations without moving from their computer. (2) Exchange of ideas. Students can have a "conversation" with experts in a particular field. (3) Comfortable communication. E-mail allows students in a variety of locations for ideas. They can "talk" to each other at various times and respond to it based on their own convenience. The exchange of ideas is kept secret. (4) Low cost. The cost of hardware, software, and internet service is nominal and continues to decline.

❖ Limitations

Incompatible material. One concern is that some topics covered on computer networks, especially on the internet, are not suitable for students. (1) Copyright. Because information is so easily accessible, it is easy for individuals to easily download a file and illegally exploit their own interests. Thus, students may write papers or projects that are simply "co-fitting" (copy paste) and not their own work. (2) Information search estimated that thousands of sites, several thousand new internet sites were added to the internet every day. This growth made finding

information more difficult. To assist in obtaining information, some commercial companies and universities provide search engines that follow web links and display the results according to your search. (3) Support without good technical support and wise management, a network of computers may “shut down” quickly. Problems with a network can paralyze a lab or even shut down an entire school. (4) Access through a wired or wireless device system or a modem, all users must be connected to the network. The computer must have the appropriate software and hardware installed in order to access the internet. In a wireless system, the user must have the necessary permissions to use the system. Signal strength is also an issue in wireless connections. The weaker the signal, the less likely it is that you will be able to send and receive data. (5) Access agreements. Another limitation is the speed at which users can access information. Extremely long waiting times can be prevented by designing compact Web pages. If you are creating your own Web page, you should take the time to combine images, colors, and text that will be meaningful to your target users. (6) Lack of quality control. Students must be critical readers and thinkers who know how to evaluate information. What is posted on the internet is not a "fact". Anyone can post anything on the Web, including information that is unnecessary, wrong or untrustworthy.(**Yuhanna, I. ., Alexander, A. ., & Kachik, A. . (2020).**)

The Random Forest algorithm is well-known for its strength, precision, and flexibility in both classification and regression applications. Nonetheless, when utilized on extensive datasets, it reveals several shortcomings that can obstruct its efficiency and practicality. This paper examines the main limitations of employing Random Forest for large-scale data analysis, such as significant computational and memory demands, prolonged training and prediction times, decreased interpretability, and challenges in managing sparse or high-dimensional data. Furthermore, the ensemble nature of the algorithm creates difficulties for real-time inference and presents constraints in extrapolation tasks. By shedding light on these challenges, this research intends to educate data scientists and engineers about the considerations necessary when implementing Random Forest in large-data contexts and proposes pathways for alternative strategies or enhancements.(**Thirupathi Kandadi**)

❖ References

1. Priyanka and Dharmender Kumar
Published Online: April 30, 2020pp 246-269
<https://doi.org/10.1504/IJIDS.2020.108141>
2. LI Kang-le, REN Zhi-lei, ZHOU Zhi-de, JIANG He

Editorial office of Computer Science; Ji suan ji ke xue, 2022-11, Vol 49 (11), p.30-38

3. Random Forests Chapter First Online: 11 September 2020 pp 33–55 Cite this chapter
4. Yuhanna, I. ., Alexander, A. ., & Kachik, A. . (2020). Advantages and disadvantages of Online Learning. *Journal Educational Verkenning*, 1(2), 13-19.
<https://doi.org/10.48173/jev.v1i2.54>
5. 16 Pages Posted: 7 May 2025 **Thirupathi Kandadi** Chaitanya Deemed to be University Date Written: April 30, 2025