

Global Refinement of Random Forest

Shaoqing Ren¹, Xudong Cao², Yichen Wei², Jian Sun²

¹University of Science and Technology of China. ²Microsoft Research.

Abstract

Random forest is well known as one of the best learning methods. In spite of its great success, it also has certain drawbacks: the heuristic learning rule does not effectively minimize the global training loss; the model size is usually too large for many real applications. To address the issues, we propose two techniques, global refinement and global pruning, to improve a pre-trained random forest. The proposed global refinement jointly relearns the leaf nodes of all trees under a global objective function so that the complementary information between multiple trees is well exploited. In this way, the fitting power of the forest is significantly enhanced. The global pruning is developed to reduce the model size as well as the over-fitting risk. The refined model has better performance and smaller storage cost, as verified in extensive experiments.

Random forest [2] is one of the most popular learning methods and has many ideal properties: 1) it is simple to understand and implement; 2) it is strong in handling non-linearity and outliers; 3) it is friendly to parallel training and large data; and 4) it is fast in testing. Recently, it has proven extremely successful on important applications in data mining [12] and computer vision [6, 11].

In spite of its great success, random forest has certain insufficiency from both theoretical and practical viewpoints. Theoretically, the heuristic learning of random forest is suboptimal in terms of minimizing training error. Specifically, each individual tree is learnt independently and greedily. Such learning does not fully utilize complementary information among different trees.

Practically, for complex real problems [3, 4, 6, 11], deep trees are usually required to fit the training data well. This results in high storage cost, which is a serious issue especially for embedded devices such as mobile phone or Kinect. While tree pruning can reduce the tree size, existing methods [7, 8] are independently performed on individual trees and could degrade the performance of random forest.

To address the above problems, we propose a simple and effective method to refine a pre-trained random forest. We notice that the learning and prediction of random forest is inconsistent: the learning of individual trees is independent but the prediction averages all trees' outputs. The loss functions implied from these two processes are actually different. This limits the fitting power of random forest. To alleviate such inconsistency, we discard the old values stored in all tree leaves of a pre-trained random forest and relearn them through a "global" refinement: *all tree leaves are simultaneously optimized by explicitly minimizing a global loss function defined on all training samples, according to the averaging prediction rule of random forest.*

The proposed global refinement can be efficiently solved with a linear support vector classification/regression machine. As a result, the complementary information between trees is exploited and the fitting power is significantly improved.

The global optimization in training might cause over-fitting for a large number of deep trees (therefore a huge number of tree leaves). To reduce the risk of over-fitting as well as the model size, we propose a global pruning method which alternates between refining all tree leaves and merging the insignificant leaves, similar to the sparse approximation [1]. In this way, the model size is significantly reduced and the generalization capability is usually improved. A toy example is shown in Figure 1.

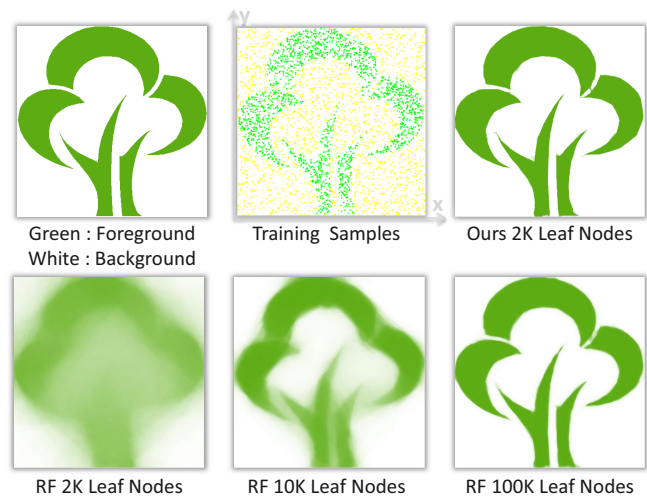


Figure 1: A toy classification task. From left to right, the first row shows the ground truth label map, training data points, and the probability map predicted by our refined forest. The second row shows the probability map predicted by standard random forest (RF) trained with various depth ($D_{\max} = 5, 8, 12$). Our refined forest can clearly separate the two classes using a smaller number of leaf nodes. In all cases, we use 100 trees in the forest.

In our various experiments, the improved random forest achieves better accuracy and smaller model size, compared to standard random forest and some state-of-the-art variants. The strong results verify the effectiveness and practicability of our approach.

Our formulation is also applicable to other ensemble tree-based models. The preliminary results on boosting trees [5] and alternating decision forests [9, 10] are encouraging.

- [1] Thomas Blumensath and Mike E Davies. Iterative thresholding for sparse approximations. *Journal of Fourier Analysis and Applications*, 14(5-6):629–654, 2008.
- [2] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.
- [3] Piotr Dollár and C Lawrence Zitnick. Structured forests for fast edge detection. In *ICCV*. IEEE, 2013.
- [4] Gabriele Fanelli, Juergen Gall, and Luc Van Gool. Real time head pose estimation with random regression forests. In *CVPR*. IEEE, 2011.
- [5] Jerome H Friedman. Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, pages 1189–1232, 2001.
- [6] Juergen Gall and Victor Lempitsky. Class-specific hough forests for object detection. In *Decision Forests for Computer Vision and Medical Image Analysis*, pages 143–157. Springer, 2013.
- [7] Manish Mehta, Jorma Rissanen, Rakesh Agrawal, et al. Mdl-based decision tree pruning. In *KDD*, 1995.
- [8] J. Ross Quinlan. Induction of decision trees. *Machine learning*, 1(1):81–106, 1986.
- [9] Samuel Schulter, Christian Leistner, Paul Wohlhart, Peter M. Roth, and Horst Bischof. Alternating regression forests for object detection and pose estimation. In *ICCV*, 2013.
- [10] Samuel Schulter, Paul Wohlhart, Christian Leistner, Amir Saffari, Peter M. Roth, and Horst Bischof. Alternating decision forests. In *CVPR*, 2013.
- [11] Jamie Shotton, Andrew Fitzgibbon, Cook Mat, Toby Sharp, Mark Finocchio, Richard Moore, Alex Kipman, and Andrew Blake. Real-time human pose recognition in parts from single depth images. In *CVPR*, 2011.
- [12] Antanas Verikas, Adas Gelzinis, and Marija Bacauskiene. Mining data with random forests: A survey and results of new tests. *Pattern Recognition*, 44(2): 330–349, 2011.