Decision Forests

A Unified Framework for Classification, Regression, Density Estimation, Manifold Learning and Semi-Supervised Learning
Decision Forests
A Unified Framework for Classification, Regression, Density Estimation, Manifold Learning and Semi-Supervised Learning

Antonio Criminisi
Microsoft Research Ltd.
Cambridge, UK
antcrim@microsoft.com

Jamie Shotton
Microsoft Research Ltd.
Cambridge, UK
jamiesho@microsoft.com

Ender Konukoglu
Microsoft Research Ltd.
Cambridge, UK
enderk@microsoft.com

Full text available at: http://dx.doi.org/10.1561/0600000035
Editorial Scope

Foundations and Trends® in Computer Graphics and Vision
will publish survey and tutorial articles in the following topics:

- Rendering: Lighting models; Forward rendering; Inverse rendering; Image-based rendering; Non-photorealistic rendering; Graphics hardware; Visibility computation
- Shape: Surface reconstruction; Range imaging; Geometric modelling; Parameterization;
- Mesh simplification
- Animation: Motion capture and processing; Physics-based modelling; Character animation
- Sensors and sensing
- Image restoration and enhancement
- Segmentation and grouping
- Feature detection and selection
- Color processing
- Texture analysis and synthesis
- Illumination and reflectance modeling
- Shape Representation
- Tracking
- Calibration
- Structure from motion
- Motion estimation and registration
- Stereo matching and reconstruction
- 3D reconstruction and image-based modeling
- Learning and statistical methods
- Appearance-based matching
- Object and scene recognition
- Face detection and recognition
- Activity and gesture recognition
- Image and Video Retrieval
- Video analysis and event recognition
- Medical Image Analysis
- Robot Localization and Navigation

Information for Librarians
Foundations and Trends® in Computer Graphics and Vision, 2011, Volume 7, 4 issues. ISSN paper version 1572-2740. ISSN online version 1572-2759. Also available as a combined paper and online subscription.
Decision Forests: A Unified Framework for Classification, Regression, Density Estimation, Manifold Learning and Semi-Supervised Learning

Antonio Criminisi¹, Jamie Shotton², and Ender Konukoglu³

¹ Microsoft Research Ltd., 7 J J Thomson Ave, Cambridge, CB3 0FB, UK, antcrim@microsoft.com
² Microsoft Research Ltd., 7 J J Thomson Ave, Cambridge, CB3 0FB, UK, jamiesho@microsoft.com
³ Microsoft Research Ltd., 7 J J Thomson Ave, Cambridge, CB3 0FB, UK, enderk@microsoft.com

Abstract

This review presents a unified, efficient model of random decision forests which can be applied to a number of machine learning, computer vision, and medical image analysis tasks.

Our model extends existing forest-based techniques as it unifies classification, regression, density estimation, manifold learning, semi-supervised learning, and active learning under the same decision forest framework. This gives us the opportunity to write and optimize the core implementation only once, with application to many diverse tasks.
The proposed model may be used both in a discriminative or generative way and may be applied to discrete or continuous, labeled or unlabeled data.

The main contributions of this review are: (1) Proposing a unified, probabilistic and efficient model for a variety of learning tasks; (2) Demonstrating margin-maximizing properties of classification forests; (3) Discussing probabilistic regression forests in comparison with other nonlinear regression algorithms; (4) Introducing density forests for estimating probability density functions; (5) Proposing an efficient algorithm for sampling from a density forest; (6) Introducing manifold forests for nonlinear dimensionality reduction; (7) Proposing new algorithms for transductive learning and active learning. Finally, we discuss how alternatives such as random ferns and extremely randomized trees stem from our more general forest model.

This document is directed at both students who wish to learn the basics of decision forests, as well as researchers interested in the new contributions. It presents both fundamental and novel concepts in a structured way, with many illustrative examples and real-world applications. Thorough comparisons with state-of-the-art algorithms such as support vector machines, boosting and Gaussian processes are presented and relative advantages and disadvantages discussed. The many synthetic examples and existing commercial applications demonstrate the validity of the proposed model and its flexibility.
Contents

1 Overview and Scope 1
1.1 A Chronological Literature Review 2

2 The Random Decision Forest Model 5
2.1 Decision Tree Basics 6
2.2 Mathematical Notation and Basic Definitions 9
2.3 Randomly Trained Decision Trees 11
2.4 Ensembles of Trees (Decision Forest) 20

3 Classification Forests 25
3.1 Classification Algorithms in the Literature 26
3.2 Specializing the Decision Forest Model for Classification 26
3.3 Effect of Model Parameters 30
3.4 Maximum-margin Properties 38
3.5 Comparisons with Alternative Algorithms 46
3.6 Human Body Tracking in Microsoft Kinect for XBox 360 48

4 Regression Forests 51
4.1 Nonlinear Regression in the Literature 51
4.2 Specializing the Decision Forest Model for Regression 52
4.3 Effect of Model Parameters 57
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.4</td>
<td>Comparison with Alternative Algorithms</td>
<td>61</td>
</tr>
<tr>
<td>4.5</td>
<td>Semantic Parsing of 3D Computed Tomography Scans</td>
<td>63</td>
</tr>
<tr>
<td>5</td>
<td>Density Forests</td>
<td>69</td>
</tr>
<tr>
<td>5.1</td>
<td>Literature on Density Estimation</td>
<td>70</td>
</tr>
<tr>
<td>5.2</td>
<td>Specializing the Forest Model for Density Estimation</td>
<td>70</td>
</tr>
<tr>
<td>5.3</td>
<td>Effect of Model Parameters</td>
<td>76</td>
</tr>
<tr>
<td>5.4</td>
<td>Comparison with Alternative Algorithms</td>
<td>79</td>
</tr>
<tr>
<td>5.5</td>
<td>Sampling from the Generative Model</td>
<td>83</td>
</tr>
<tr>
<td>5.6</td>
<td>Dealing with Non-function Relations</td>
<td>85</td>
</tr>
<tr>
<td>5.7</td>
<td>Quantitative Analysis</td>
<td>91</td>
</tr>
<tr>
<td>6</td>
<td>Manifold Forests</td>
<td>95</td>
</tr>
<tr>
<td>6.1</td>
<td>Literature on Manifold Learning</td>
<td>96</td>
</tr>
<tr>
<td>6.2</td>
<td>Specializing the Forest Model for Manifold Learning</td>
<td>97</td>
</tr>
<tr>
<td>6.3</td>
<td>Experiments and the Effect of Model Parameters</td>
<td>104</td>
</tr>
<tr>
<td>6.4</td>
<td>Learning Manifold of Object Shapes</td>
<td>110</td>
</tr>
<tr>
<td>6.5</td>
<td>Learning Manifold of Text Documents</td>
<td>112</td>
</tr>
<tr>
<td>6.6</td>
<td>Discussion</td>
<td>113</td>
</tr>
<tr>
<td>7</td>
<td>Semi-supervised Forests</td>
<td>115</td>
</tr>
<tr>
<td>7.1</td>
<td>Literature on Semi-supervised Learning</td>
<td>116</td>
</tr>
<tr>
<td>7.2</td>
<td>Specializing the Decision Forest Model</td>
<td>117</td>
</tr>
<tr>
<td></td>
<td>for Semi-supervised Classification</td>
<td></td>
</tr>
<tr>
<td>7.3</td>
<td>Label Propagation in Transduction Forest</td>
<td>119</td>
</tr>
<tr>
<td>7.4</td>
<td>Induction from Transduction</td>
<td>121</td>
</tr>
<tr>
<td>7.5</td>
<td>Examples, Comparisons and Effect</td>
<td>123</td>
</tr>
<tr>
<td></td>
<td>of Model Parameters</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Random Ferns and Other Forest Variants</td>
<td>129</td>
</tr>
<tr>
<td>8.1</td>
<td>Extremely Randomized Trees</td>
<td>129</td>
</tr>
<tr>
<td>8.2</td>
<td>Random Ferns</td>
<td>130</td>
</tr>
<tr>
<td>8.3</td>
<td>Online Forest Training</td>
<td>132</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
<td></td>
</tr>
<tr>
<td>--------------------------------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>8.4 Structured-output Forests</td>
<td>132</td>
<td></td>
</tr>
<tr>
<td>8.5 Further Forest Variants</td>
<td>134</td>
<td></td>
</tr>
<tr>
<td>9 Conclusions</td>
<td>135</td>
<td></td>
</tr>
<tr>
<td>Appendix A — Deriving the Regression Information Gain</td>
<td>137</td>
<td></td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>141</td>
<td></td>
</tr>
<tr>
<td>References</td>
<td>143</td>
<td></td>
</tr>
</tbody>
</table>
Overview and Scope

This review presents a unified, efficient model of random decision forests which can be used in a number of applications such as scene recognition from photographs, object recognition in images, automatic diagnosis from radiological scans and semantic text parsing. Such applications have traditionally been addressed by different, supervised or unsupervised machine learning techniques.

In this review, we formulate diverse learning tasks such as regression, classification and semi-supervised learning as instances of the same general decision forest model. The unified framework further extends to novel uses of forests in tasks such as density estimation and manifold learning. The underlying unified framework gives us the opportunity to implement and optimize the general algorithm for all these tasks only once, and then adapt it to individual applications with relatively small changes.

This review is directed at engineers and PhD students who wish to learn the basics of decision forests as well as more senior researchers interested in the new research contributions.

We begin by presenting a roughly chronological, non-exhaustive survey of decision trees and forests, and their use in the past two decades. Further references will be available in the relevant sections.
Overview and Scope

1.1 A Chronological Literature Review

One of the earlier works on decision trees is the seminal “Classification and Regression Trees (CART)” book by Breiman et al. [12], where the authors describe the basics of decision trees and their use for both classification and regression problems. Following that publication researchers then focused on algorithms for constructing (learning) optimal decision trees for different tasks using available training data. For this purpose, one of the most popular algorithms is “C4.5” of Quinlan [81]. Although, decision trees were proven to be useful, their application remained limited to relatively low dimensional data.

In the nineties, researchers discovered how using ensembles of learners (e.g., generic “weak” classifiers) yields greater accuracy and generalization. This seems particularly true for high dimensional data, as often encountered in real life applications. One of the earliest references to ensemble methods is the boosting algorithm of Schapire [87], where the author discusses how iterative re-weighting of training data can be used to build “strong” classifiers as linear combination of many weak ones.

Combining the ideas of decision trees and ensemble methods gave rise to decision forests, that is, ensembles of randomly trained decision trees. The idea of constructing and using ensembles of trees with randomly generated node tests was introduced for the first time in the work of Amit and Geman [1, 2] for handwritten digit recognition. In that work the authors also propose using the mean of the tree probabilities as output of the tree ensemble.

In the subsequent work of Ho [47] tree training via randomized partitioning of the feature space is discussed further, and in [48] forests are shown to yield superior generalization to both boosting and pruned C4.5-trained trees, on some tasks. The author also shows comparisons between different split functions in the tree nodes.

Breiman’s later work in [10, 11] further consolidated the role of random forests and popularized their use. There, the author introduces a different way of injecting randomness in the forest by randomly sampling the labeled training data (namely “bagging”). The author
also describes techniques for predicting the forest test error based on measures of tree strength and correlation.

In computer vision, ensemble methods became popular with the seminal face and pedestrian detection papers of Viola and Jones [107, 108]. Random decision forests were used in [63] for image classification and in [60] for keypoint tracking in videos. Recent years have seen an explosion of forest-based techniques in the machine learning, vision and medical imaging literature [9, 15, 25, 35, 37, 58, 59, 65, 67, 68, 69, 74, 79, 89, 92, 97, 110]. Decision forests compare favorably with respect to other techniques [15] and have lead to one of the biggest success stories of computer vision in recent years: the Microsoft Kinect for XBox 360 [39, 91, 66].
References


References


References


References


References


References


