

Decision tree pruning

Pruning is a technique in machine learning and search algorithms that reduces the size of decision trees by removing sections of the tree that provide little power to classify instances. Pruning reduces the complexity of the final classifier, and hence improves predictive accuracy by the reduction of overfitting.

Contents

Introduction

Techniques

- Reduced error pruning

- Cost complexity pruning

See also

References

Further reading

External links

Introduction

One of the questions that arises in a decision tree algorithm is the optimal size of the final tree. A tree that is too large risks overfitting the training data and poorly generalizing to new samples. A small tree might not capture important structural information about the sample space. However, it is hard to tell when a tree algorithm should stop because it is impossible to tell if the addition of a single extra node will dramatically decrease error. This problem is known as the horizon effect. A common strategy is to grow the tree until each node contains a small number of instances then use pruning to remove nodes that do not provide additional information.^[1]

Pruning should reduce the size of a learning tree without reducing predictive accuracy as measured by a cross-validation set. There are many techniques for tree pruning that differ in the measurement that is used to optimize performance.

Techniques

Pruning can occur in a top down or bottom up fashion. A top down pruning will traverse nodes and trim subtrees starting at the root, while a bottom up pruning will start at the leaf nodes. Below are two popular pruning algorithms.

Reduced error pruning

One of the simplest forms of pruning is reduced error pruning. Starting at the leaves, each node is replaced with its most popular class. If the prediction accuracy is not affected then the change is kept. While somewhat naive, reduced error pruning has the advantage of **simplicity and speed**.

Cost complexity pruning

Cost complexity pruning generates a series of trees $T_0 \dots T_m$ where T_0 is the initial tree and T_m is the root alone. At step i , the tree is created by removing a subtree from tree $i - 1$ and replacing it with a leaf node with value chosen as in the tree building algorithm. The subtree that is removed is chosen as follows:

1. Define the error rate of tree T over data set S as $\text{err}(T, S)$.
2. The subtree that minimizes $\frac{\text{err}(\text{prune}(T, t), S) - \text{err}(T, S)}{|\text{leaves}(T)| - |\text{leaves}(\text{prune}(T, t))|}$ is chosen for removal.

The function $\text{prune}(T, t)$ defines the tree obtained by pruning the subtrees t from the tree T . Once the series of trees has been created, the best tree is chosen by generalized accuracy as measured by a training set or cross-validation.

See also

- [Alpha-beta pruning](#)
- [Artificial neural network](#)
- [Null-move heuristic](#)

References

- [Judea Pearl, *Heuristics*, Addison-Wesley, 1984](#)
- [Pessimistic Decision tree pruning based on Tree size^{\[2\]}](#)
- 1. Trevor Hastie, Robert Tibshirani, and Jerome Friedman. The Elements of Statistical Learning. Springer: 2001, pp. 269-272
- 2. Mansour, Y. (1997), "Pessimistic decision tree pruning based on tree size" (<http://citeseer.ist.psu.edu/76752.html>), *Proc. 14th International Conference on Machine Learning*. 195–201

Further reading

- [MDL based decision tree pruning \(http://web.cs.iastate.edu/~honavar/kdd95_md1.pdf\)](http://web.cs.iastate.edu/~honavar/kdd95_md1.pdf)
- [Decision tree pruning using backpropagation neural networks \(http://www.cp.eng.chula.ac.th/~boonserm/publication/ijcnn_kc2001.pdf\)](http://www.cp.eng.chula.ac.th/~boonserm/publication/ijcnn_kc2001.pdf)

External links

- [Fast, Bottom-Up Decision Tree Pruning Algorithm \(http://www.cis.upenn.edu/~mkearns/papers/pruning.pdf\)](http://www.cis.upenn.edu/~mkearns/papers/pruning.pdf)
- [Introduction to Decision tree pruning \(http://www.math.tau.ac.il/~mansour/ml-course/scribe11.ps\)](http://www.math.tau.ac.il/~mansour/ml-course/scribe11.ps)

Retrieved from "https://en.wikipedia.org/w/index.php?title=Decision_tree_pruning&oldid=946949430"

Text is available under the [Creative Commons Attribution-ShareAlike License](#); additional terms may apply. By using this site, you agree to the [Terms of Use and Privacy Policy](#). Wikipedia® is a registered trademark of the [Wikimedia Foundation, Inc.](#), a non-profit organization.