Exploring the Impact of Decision Tree Depth

Alic Szecsei
University of Iowa
alic-szecsei@uiowa.edu

Diego Castaneda
University of Iowa
diego-castaneda@uiowa.edu

Willem DeJong
University of Iowa
willem-dejong@uiowa.edu

ABSTRACT

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Morbi malesuada, quam in pulvinar varius, metus nunc fermentum urna, id sollicitudin purus odio sit amet enim. Aliquam ullamcorper eu ipsum vel mollis. Curabitur quis dictum nisl. Phasellus vel semper risus, et lacinia dolor. Integer ultricies commodo sem nec semper.

CCS CONCEPTS

• **Computing methodologies → Machine learning**; *Supervised learning by classification*; *Classification and regression trees*; Cross-validation.

KEYWORDS

decision trees, model selection

1 BACKGROUND AND MOTIVATION

Construction of decision trees commonly occurs in two phases: first, a "growing" phase, in which data is used to expand the decision tree, followed by a "pruning" phase, in which noisy or otherwise meaningless nodes are removed from the tree and replaced with leaves. This second phase is used to combat overfitting and eliminate noise.

Additionally, decision tree models can be constrained by size to combat overfitting. Russell and Norvig [3] showcase an implementation of restricting a decision tree to be beneath a maximum size by generating the tree in breadth-first fashion, and stopping when the maximum number of nodes has been reached. Garofalakis and Hyun [1] support this type of algorithm, rather than implementing the constraint in the pruning phase: after all, if the program knows a branch of a tree will later be pruned, there is no purpose in constructing that tree branch.

2 METHODS

The foundation of our decision tree algorithm was the one provided by Russell and Norvig [3] which, in turn, is based on the ID3 algorithm [2].

We selected four data sets from the Machine Learning Repository at http://archive.ics.uci.edu/ml/datasets.php. We then ran each through the DecisionTreeLearning algorithm with maximum depths ranging from 1 to 10 and performed k-fold cross-validation to determine error, with $k = 4$.

3 RESULTS

4 DISCUSSION

ACKNOWLEDGMENTS

This work was supported in part by the University of Iowa.

Function DecisionTreeLearning(*examples*, *attributes*, *parent_examples*, *depth*)

```
if examples is empty then
    return Plurality-Value(parent_examples)
else if depth = 0 then
    return Plurality-Value(examples)
else if all examples have the same classification c then
    return a leaf node c
else if attributes is empty then
    return Plurality-Value(examples)
end
A ← argmax_{a∈attributes} Importance(a, examples);
tree ← a new decision tree with root test A;
foreach value v_k of A do
    exs ← {e : e ∈ examples and e.A = v_k};
    subtree ←
     DecisionTreeLearning(exs, attributes - A, examples, depth - 1);
    add a branch to tree with label (A = v_k) and subtree
     subtree;
end
return tree
```

REFERENCES

[1] Minos Garofalakis, Dongjoon Hyun, Rajeev Rastogi, and Kyuseok Shim. 2000. Efficient Algorithms for Constructing Decision Trees with Constraints. In *Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '00)*. ACM, New York, NY, USA, 335–339. https://doi.org/10.1145/347090.347163

[2] J. R. Quinlan. 1986. Induction of decision trees. *Machine Learning* 1, 1 (01 Mar 1986), 81–106. https://doi.org/10.1007/BF00116251

[3] Stuart J. Russell and Peter Norvig. 2010. *Artificial intelligence: a modern approach* (3rd ed.). Prentice Hall.

A RESEARCH METHODS

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Morbi malesuada, quam in pulvinar varius, metus nunc fermentum urna, id sollicitudin purus odio sit amet enim. Aliquam ullamcorper eu ipsum vel mollis. Curabitur quis dictum nisl. Phasellus vel semper risus, et lacinia dolor. Integer ultricies commodo sem nec semper.

B ONLINE RESOURCES

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Morbi malesuada, quam in pulvinar varius, metus nunc fermentum urna, id sollicitudin purus odio sit amet enim. Aliquam ullamcorper eu ipsum vel mollis. Curabitur quis dictum nisl. Phasellus vel semper risus, et lacinia dolor. Integer ultricies commodo sem nec semper.