Group Study

Rule Post Pruning

Introduction

In machine learning, particularly in decision tree learning, overfitting is a common problem where a model fits the training data too well, capturing noise and reducing its ability to generalize to unseen data. Rule post-pruning is an effective technique to address this by simplifying the model after it has been trained. Rather than halting tree growth early (pre-pruning), rule post-pruning allows the tree to fully grow and then trims it back, improving its generalization performance. This method has been found to yield highly accurate hypotheses and is widely used in practical applications of decision trees.

Objective:

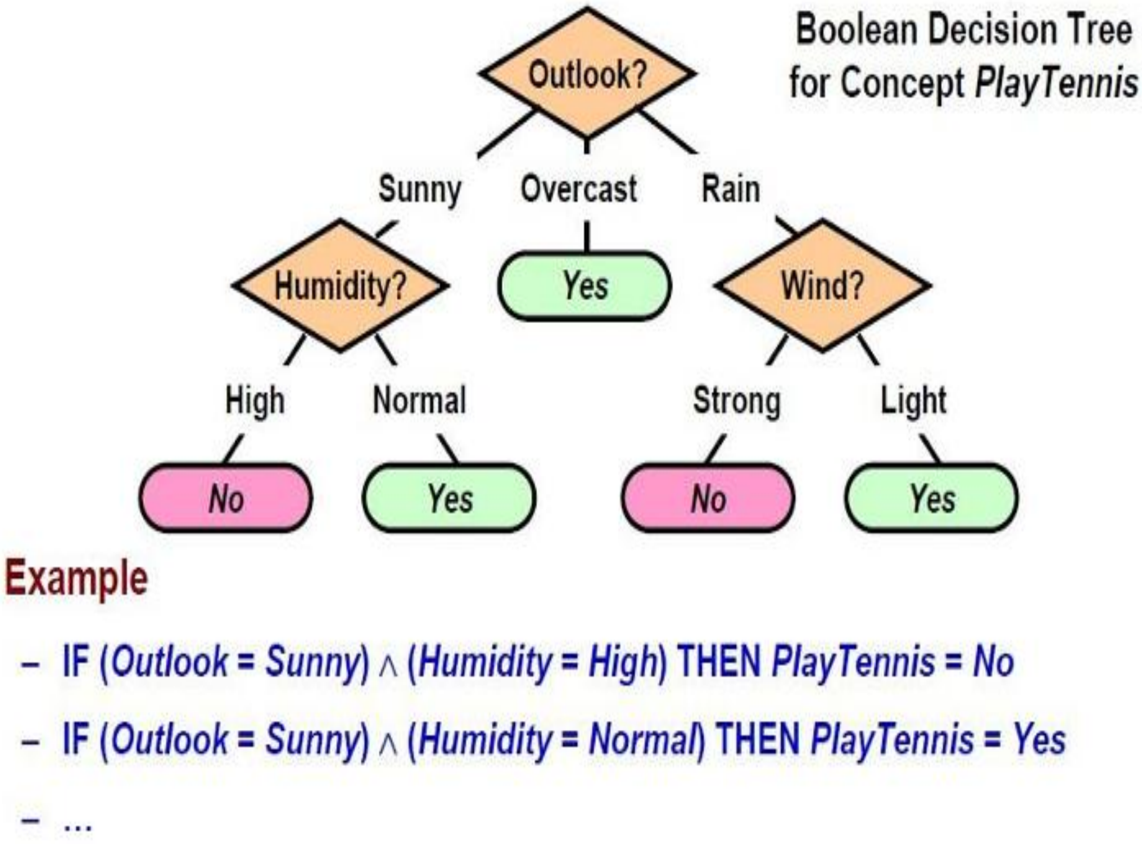
The objective of this study is to explore and understand the concept of rule post-pruning in decision tree learning, focusing on its methodology, effectiveness in reducing overfitting, and its application in generating high-accuracy, generalizable, and interpretable hypotheses. This includes analysing the step-by-step process of converting a fully grown decision tree into a simplified rule-based model and evaluating its usefulness across various real-world domains.

Process:

Rule post-pruning is successful method for finding high accuracy hypotheses

- Rule post-pruning involves the following steps:
- Infer the decision tree from the training set, growing the tree until the training data is fit as well as possible and allowing overfitting to occur.
- Convert the learned tree into an equivalent set of rules by creating one rule for each path from the root node to a leaf node.
- Prune(generalize) each rule by removing any preconditions that result in improving its estimated accuracy.
- Sort the pruned rules by their estimated accuracy, and consider them in this sequence when classifying subsequent instances

Converting a Decision Tree into Rules



For example, consider the decision tree. The leftmost path of the tree in below figure is translated into the rule.

IF (Outlook = Sunny) ^ (Humidity = High)

THEN Play Tennis = No

Given the above rule, rule post-pruning would consider removing the preconditions

(Outlook = Sunny) and (Humidity = High)

- It would select whichever of these pruning steps produced the greatest improvement in estimated rule accuracy, then consider pruning the second precondition as a further pruning step.

- No pruning step is performed if it reduces the estimated rule accuracy

There are three main advantages by converting the decision tree to rules before pruning:

- Allows separate pruning decisions for the same attribute in different paths.
- Avoids complex restructuring of the tree when pruning high-level nodes.
- Improves readability and makes the model easier to understand.

Applications of Rule Post-Pruning

- **Medical Diagnosis:** Helps simplify complex diagnostic rules for better interpretability while maintaining accuracy**.**

- **Finance:** Used in credit scoring and risk assessment to reduce overfitting while keeping decision criteria understandable.

- **Customer Behaviour Prediction:** Applied in marketing and recommendation systems to extract general patterns from customer data.

- **Fraud Detection:** Detects anomalous behaviours by refining rules that distinguish between legitimate and fraudulent activities.

Conclusion:

Rule post-pruning enhances the generalization ability of decision tree models by simplifying rules without compromising their predictive accuracy. As illustrated in the example, the process involves evaluating each precondition in a rule and selectively removing those that do not contribute to or improve accuracy. By pruning only when it results in a performance gain, the method ensures that the final rules remain both effective and interpretable. This selective simplification helps avoid overfitting and leads to more robust models suitable for real-world decision-making tasks**.**

Presented by:

K. Swathi (22491A4722)

Sd. Khadharunnisa (22491A4746)

Y. Radha Krishna (22491A4756)