

Innovative Teaching Report

Academic Year: 2023 – 2024

Subject: CST466 Data Mining

Faculty: Mrs. Ansia S

Innovative Teaching Method: Flipped Classroom

Topic/Question: Previous year university question on decision tree classification

Q) Explain the working of a decision tree. Use ID3 algorithm and find the best attribute to split at the root level of a decision tree.

Age	Competition	Type	Class (profit)
Old	Yes	Software	Down
Old	No	Software	Down
Old	No	Hardware	Down
Mid	Yes	Software	Down
Mid	Yes	Hardware	Down
Mid	No	Hardware	Up
Mid	No	Software	Up
New	Yes	Software	Up
New	No	Hardware	Up
New	No	Software	Up

Introduction

In the academic year 2023-2024, an innovative teaching method known as the Flipped Classroom was employed for the subject CST466 Data Mining. The flipped classroom approach was utilized to enhance student engagement and deepen understanding of the topic of decision tree classification, specifically using the ID3 algorithm.

Innovative Teaching Method: Flipped Classroom

Implementation in CST466 Data Mining

For the topic of decision tree classification using the ID3 algorithm, students were provided with pre-recorded lectures and reading materials to study before the class. These materials covered the basics of decision trees, the ID3 algorithm, and entropy calculations.

Topic Overview: Decision Tree Classification and ID3 Algorithm

Decision Tree Classification

A decision tree is a flowchart-like tree structure where each internal node represents a test on an attribute, each branch represents an outcome of the test, and each leaf node represents a class label. The paths from the root to the leaf represent classification rules.

ID3 Algorithm

The ID3 (Iterative Dichotomiser 3) algorithm is used to generate a decision tree from a dataset. It works as follows:

Calculate Entropy: Measure the impurity or uncertainty in the dataset.

Information Gain: For each attribute, calculate the information gain when splitting on that attribute.

Select Best Attribute: Choose the attribute with the highest information gain to split the dataset.

Recursive Splitting: Apply the same process recursively to the subsets of the dataset.

Problem Statement

Find the best attribute to split at the root level of a decision tree using the ID3 algorithm.

Given Dataset:

Age	Competition	Type	Class (profit)
Old	Yes	Software	Down
Old	No	Software	Down
Old	No	Hardware	Down
Mid	Yes	Software	Down
Mid	Yes	Hardware	Down
Mid	No	Hardware	Up
Mid	No	Software	Up

New Yes Software Up

New No Hardware Up

New No Software Up

Calculation Using ID3 Algorithm

Step 1: Calculate Entropy of the Dataset

$$E(S) = -p(Up) \log_2(p(Up)) - p(Down) \log_2(p(Down))$$

There are 5 "Up" and 5 "Down" outcomes.

$$E(S) = -\frac{5}{10} \log_2\left(\frac{5}{10}\right) - \frac{5}{10} \log_2\left(\frac{5}{10}\right)$$

$$E(S) = -0.5 \log_2(0.5) - 0.5 \log_2(0.5)$$

$$E(S) = -0.5 \times -1 - 0.5 \times -1$$

$$E(S) = 1$$

Step 2: Calculate Information Gain for Each Attribute

Information Gain for Age

Split on "Old", "Mid", "New"

$$E(S_{Old}) = -3/3 \log_2(3/3) - 0/3 \log_2(0/3) = 0$$

$$E(S_{Mid}) = -2/4 \log_2(2/4) - 2/4 \log_2(2/4) = 1$$

$$E(S_{New}) = -0/3 \log_2(0/3) - 3/3 \log_2(3/3) = 0$$

$$E(S|Age) = (3/10) \cdot 0 + (4/10) \cdot 1 + (3/10) \cdot 0 = 0.4$$

$$IG(S, Age) = E(S) - E(S|Age) = 1 - 0.4 = 0.6$$

Information Gain for Competition

Split on "Yes", "No"

$$E(S_{Yes}) = -2/4 \log_2(2/4) - 2/4 \log_2(2/4) = 1$$

$$E(S_{No}) = -3/6 \log_2(3/6) - 3/6 \log_2(3/6) = 1$$

$$E(S|Competition) = (4/10) \cdot 1 + (6/10) \cdot 1 = 1$$

$$IG(S, Competition) = E(S) - E(S|Competition) = 1 - 1 = 0$$

Information Gain for Type

Split on "Software", "Hardware"

$$E(S_{Software}) = -2/5 \log_2(2/5) - 3/5 \log_2(3/5) = 0.97$$

$$E(S_{Hardware}) = -3/5 \log_2(3/5) - 2/5 \log_2(2/5) = 0.97$$

$$E(S|Type) = (5/10) \cdot 0.97 + (5/10) \cdot 0.97 = 0.97$$

$$IG(S, Type) = E(S) - E(S|Type) = 1 - 0.97 = 0.03$$

Step 3: Select the Best Attribute

The attribute with the highest information gain is Age, with an information gain of 0.6.

Classroom Activity

During the classroom session, students were asked to work in groups to verify the calculations and discuss the results. This interactive approach helped students to better understand the ID3 algorithm and its application in decision tree classification.

Conclusion

The flipped classroom method effectively engaged students and deepened their understanding of the decision tree classification using the ID3 algorithm. By allowing students to explore the topic independently before class and then applying their knowledge during the session, the learning experience was significantly enhanced.

Prepared by: Mrs. Ansia S

