Decision trees. Applications

Lecture 2.5
Decision trees for classification

- Classify and make transparent decision
- Each class leaf has its own rule path
- The same result by different reasons

Applications

- Limitations
- Real-life examples
- Extracting rules from trees
Decision trees for data exploration

- The most important attributes are at the top of the tree
- Start each data mining project from exploring the most important attributes with decision trees

• ID3 algorithm
• Design issues
  • Split criteria
  • Stop criteria
  • Multi-valued attributes
  • Numeric attributes
  • Missing values
  • Overfitting

Applications
• Limitations
• Real-life examples
• Extracting rules from trees
When (not) to use decision trees

**Good performance (use decision trees)**
- The factors of decision are not less important than the classification accuracy
- The goal is to assign each record to one of a few broad categories (Categorical attributes with low cardinality*)
- You suspect that there is a set of objective rules underlying the data

**Not that good performance (use something else)**
- Continuous numeric attributes, ordinal attributes
- Hierarchical relationships between classes
- High-cardinality attributes
- Numeric value prediction

-ID3 algorithm
- Design issues
  - Split criteria
  - Stop criteria
  - Multi-valued attributes
  - Numeric attributes
  - Missing values
  - Overfitting
- Applications
- Limitations
  - Real-life examples
  - Extracting rules from trees

*cardinality - the number of possible distinct values
Limitations. Rectilinear decision boundaries

- Boolean split: the instances are divided by the boundaries which are parallel to the axes
- Solution: use all reasonable combinations of attributes.
Non-rectilinear boundaries: attribute combinations

One-level decision tree

\[ 0.25L + R < 1.5 \]

true \rightarrow No

false \rightarrow Yes

0.25L + R < 1.5

true \rightarrow No

false \rightarrow Yes
Decision trees in real life

- Selecting the most promising eggs for in-vitro fertilization – England, 2000
- Soybean disease classification – 1979, 97% accuracy vs. 72% by human expert
- Classification system for serial criminal patterns (CSSCP) - using three years' worth of data on armed robbery, the system was able to spot 10 times as many patterns as a team of experienced detectives with access to the same data.
- Screening potential terrorists and drug smugglers at border crossings

ID3 algorithm
Design issues
  - Split criteria
  - Stop criteria
  - Multi-valued attributes
  - Numeric attributes
  - Missing values
  - Overfitting
Applications
Limitations
Real-life examples
Extracting rules from trees
Border crossing: gross oversimplification

- Age: 20-25
- Gender: male
- Nationality: Saudi Arabia
- Country of residence: Germany
- Visa status: student
- University: unknown
- # times entering the country in the past year: 3
- Countries visited during the past 3 years: U.K., Pakistan
- Flying lessons: yes

Assessment: possible terrorist (probability 29%)
Action: detain and report

Carnival Booth: An Algorithm for Defeating the Computer-Assisted Passenger Screening System
From trees to rules: how?

• How can we produce a set of rules from a decision tree?
From trees to rules – simple

- One rule for each leaf

**If** Temp = “Warm” **then** play

**If** Temp = “Hot” **then** play

**If** Temp = “Chilly” and Outlook=“Sunny” **then** play

**Default**: no play
From trees to rules – simple

- The set of rules can be minimized

If Temp = “Chilly” and (Outlook=“Sunny” or Outlook = “Overcast”) then no play
Default: play

- ID3 algorithm
- Design issues
  - Split criteria
  - Stop criteria
- Multi-valued attributes
- Numeric attributes
- Missing values
- Overfitting
- Applications
- Limitations
- Real-life examples

Extracting rules from trees
Difference between decision trees and rules

- Rules are more readable than decision trees

- Decision trees describe the **general concept** extracted from the data, while each rule represents a **nugget of knowledge**

- Trees contain predictions for **all class variables**, while each rule predicts only **one class value**

- ID3 algorithm
- Design issues
  - Split criteria
  - Stop criteria
  - Multi-valued attributes
  - Numeric attributes
  - Missing values
  - Overfitting
- Applications
- Limitations
- Real-life examples
- Extracting rules from trees