Decision tree

From Wikipedia, the free encyclopedia

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikipedia store

Interaction
Help
About Wikipedia
Community portal
Recent changes
Contact page

Tools
What links here
Related changes
Upload file
Special pages
Permanent link
Page information
Wikidata item
Cite this page

Print/export
Create a book
Download as PDF
Printable version

Languages
العربية
Български
Čeština
Deutsch
Español
فارسی
Français
한국어
Italiano
עברית
Magyar
Nederlands
日本語
Norsk bokmål
Polski
Português
Русский
ไทย
Українська
Tiếng Việt
中文
Edit links

This article **needs additional citations for verification**. Please help improve this article by adding citations to reliable sources. Unsourced material may be challenged and removed. *(October 2013)*

This article is about decision trees in decision analysis. For the use of the term in machine learning, see Decision tree learning.

A **decision tree** is a decision support tool that uses a tree-like graph or model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility. It is one way to display an algorithm.

Decision trees are commonly used in operations research, specifically in decision analysis, to help identify a strategy most likely to reach a goal.

Traditionally, decision trees have been created manually.

Contents [hide]

Overview [edit]

A decision tree is a flowchart-like structure in which each internal node represents a "test" on an attribute (e.g. whether a coin flip comes up heads or tails), each branch represents the outcome of the test and each leaf node represents a class label (decision taken after computing all attributes). The paths from root to leaf represents classification rules.

In decision analysis a decision tree and the closely related influence diagram are used as a visual and analytical decision support tool, where the expected values (or expected utility) of competing alternatives are calculated.

A decision tree consists of 3 types of nodes:

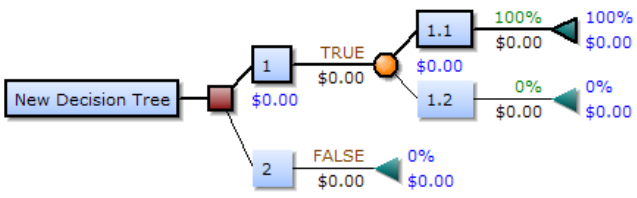
1. Decision nodes - commonly represented by squares
2. Chance nodes - represented by circles
3. End nodes - represented by triangles

Decision trees are commonly used in operations research, specifically in decision analysis, to help identify a strategy most likely to reach a goal. If in practice decisions have to be taken online with no recall under incomplete knowledge, a decision tree should be paralleled by a probability model as a best choice model or online selection model algorithm. Another use of decision trees is as a descriptive means for calculating conditional probabilities.

Decision trees, influence diagrams, utility functions, and other decision analysis tools and methods are taught to undergraduate students in schools of business, health economics, and public health, and are examples of operations research or management science methods.

Decision tree building blocks [edit]

Decision tree elements [edit]



Drawn from left to right, a decision tree has only burst nodes (splitting paths) but no sink nodes (converging paths).

Therefore, used manually, they can grow very big and are then often hard to draw fully by hand. Traditionally, decision trees have been created manually - as the aside example shows - although increasingly, specialized software is employed.

Decision rules [edit]

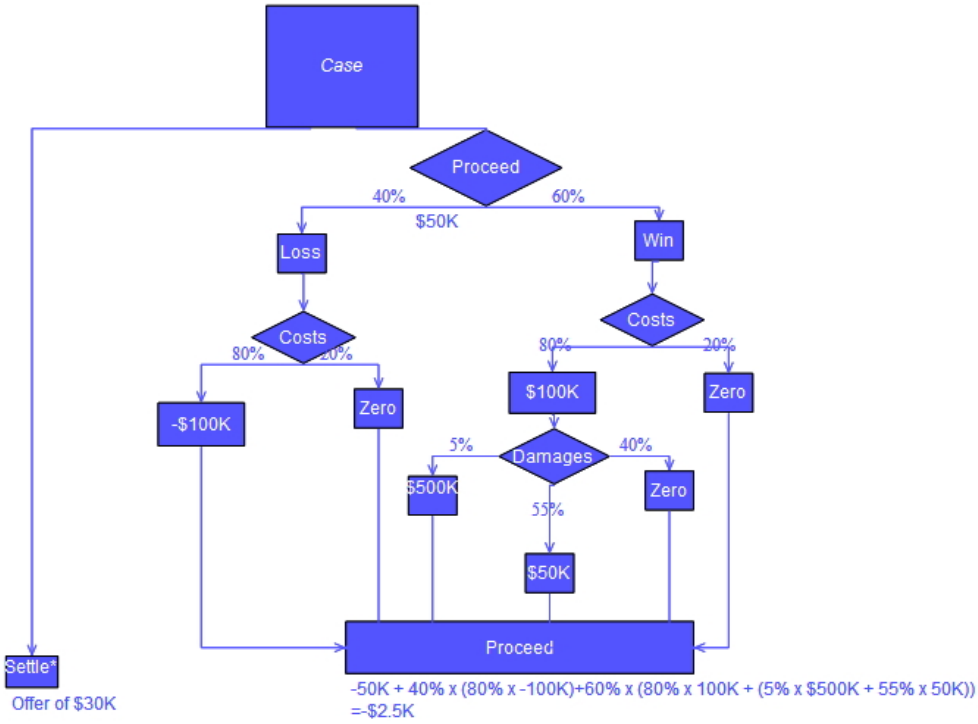
The decision tree can be linearized into **decision rules**,[1] where the outcome is the contents of the leaf node, and the conditions along the path form a conjunction in the if clause. In general, the rules have the form:

if condition1 *and* condition2 *and* condition3 *then* outcome.

Decision rules can also be generated by constructing association rules with the target variable on the right.

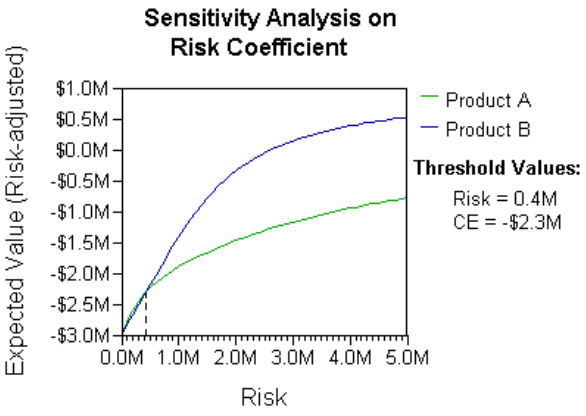
Decision tree using flowchart symbols [edit]

Commonly a decision tree is drawn using flowchart symbols as it is easier for many to read and understand.



Analysis example [edit]

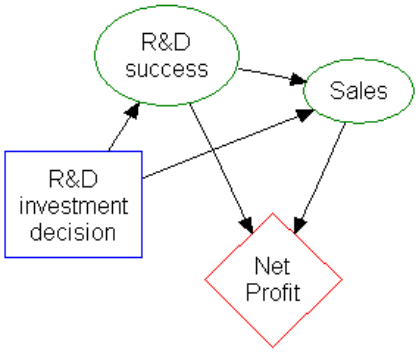
Analysis can take into account the decision maker's (e.g., the company's) preference or utility function, for example:



The basic interpretation in this situation is that the company prefers B's risk and payoffs under realistic risk preference coefficients (greater than $400K—in that range of risk aversion, the company would need to model a third strategy, "Neither A nor B").

Influence diagram [edit]

Much of the information in a decision tree can be represented more compactly as an influence diagram, focusing attention on the issues and relationships between events.



The squares represent decisions, the ovals represent action, and the diamond represents results.

Advantages and disadvantages [edit]

Among decision support tools, decision trees (and influence diagrams) have several advantages. Decision trees:

- Are simple to understand and interpret. People are able to understand decision tree models after a brief explanation.
- Have value even with little hard data. Important insights can be generated based on experts describing a situation (its alternatives, probabilities, and costs) and their preferences for outcomes.
- Allow the addition of new possible scenarios
- Help determine worst, best and expected values for different scenarios
- Use a white box model. If a given result is provided by a model.
- Can be combined with other decision techniques.

Disadvantages of decision trees:

- For data including categorical variables with different number of levels, information gain in decision trees are biased in favor of those attributes with more levels.[2]
- Calculations can get very complex particularly if many values are uncertain and/or if many outcomes are linked.

See also [edit]

- Decision cycle
- Decision tables
- Decision tree model of computation
- DRAKON
- Expectiminimax tree
- Influence diagram
- Markov chain
- Morphological analysis
- Random forest
- Odds algorithm
- Operations research
- Topological combinatorics
- Truth table

References [edit]

1. ^ Quinlan, J. R. (1987). "Simplifying decision trees". *International Journal of Man-Machine Studies* **27** (3): 221. doi:10.1016/S0020-7373(87)80053-6.
2. ^ Deng,H.; Runger, G.; Tuv, E. (2011). *Bias of importance measures for multi-valued attributes and solutions*. Proceedings of the 21st International Conference on Artificial Neural Networks (ICANN).

Further reading [edit]

- Cha, Sung-Hyuk; Tappert, Charles C (2009). "A Genetic Algorithm for Constructing Compact Binary Decision Trees". *Journal of Pattern Recognition Research* **4** (1): 1–13. doi:10.13176/11.44.

External links [edit]

- SilverDecisions: a free and open source decision tree software
- Decision Tree Analysis mindtools.com
- Decision Analysis open course at George Mason University
- Extensive Decision Tree tutorials and examples

Wikimedia Commons has media related to ***decision diagrams***.

Categories: Decision trees | Decision theory