Comprehensibility of Classification Trees – Survey Design Validation

Rok Piltaver, Mitja Luštrek, Matjaž Gams
Jozef Stefan Institute - Department of Intelligent Systems
Jozef Stefan International Postgraduate School
Sanda Martinčić - Ipšić
University of Rijeka - Department of Informatics

6th International Conference on Information Technologies and Information Society
Šmarješke toplice, Slovenia, November 6-7, 2014
Outline

• Introduction and motivation
• Survey design
• Survey implementation
• Survey design verification
Machine learning (ML) and data mining (DM)

Attributes | Class
---|---
gender | sport
height | 
weight | 

ML algorithm (e.g. C4.5)

Examples

<table>
<thead>
<tr>
<th>gender</th>
<th>height</th>
<th>weight</th>
<th>sport</th>
</tr>
</thead>
<tbody>
<tr>
<td>male</td>
<td>190</td>
<td>86.04</td>
<td>basketball</td>
</tr>
<tr>
<td>female</td>
<td>191</td>
<td>76.39</td>
<td>basketball</td>
</tr>
<tr>
<td>male</td>
<td>175</td>
<td>57.07</td>
<td>marathon</td>
</tr>
<tr>
<td>female</td>
<td>168</td>
<td>46.12</td>
<td>marathon</td>
</tr>
<tr>
<td>female</td>
<td>162</td>
<td>62.24</td>
<td>weight-lifting</td>
</tr>
<tr>
<td>male</td>
<td>194</td>
<td>87.71</td>
<td>basketball</td>
</tr>
</tbody>
</table>

Learning data set

<table>
<thead>
<tr>
<th>gender</th>
<th>height</th>
<th>weight</th>
<th>sport</th>
</tr>
</thead>
<tbody>
<tr>
<td>male</td>
<td>188</td>
<td>94.36</td>
<td>?</td>
</tr>
<tr>
<td>female</td>
<td>164</td>
<td>69.32</td>
<td>?</td>
</tr>
<tr>
<td>female</td>
<td>186</td>
<td>75.42</td>
<td>?</td>
</tr>
<tr>
<td>male</td>
<td>178</td>
<td>56.8</td>
<td>?</td>
</tr>
<tr>
<td>female</td>
<td>166</td>
<td>45.17</td>
<td>?</td>
</tr>
<tr>
<td>male</td>
<td>188</td>
<td>97.9</td>
<td>?</td>
</tr>
</tbody>
</table>

New data without class

ML Model (Classifier)

Predictions

sport

- weight-lifting
- basketball
- marathon
- weight-lifting
Comprehensibility - definition

- The ability to understand the output of induction algorithm [1].
- The ability to understand the logic behind a prediction of a model [2].
- How well humans grasp the induced classifier [3].

Also referred to as:
- interpretability [4] or understandability [5].

Related to but not the same as:
- transparency, complexity, readability.
Comprehensibility - examples

(fins = yes) and (eggs = yes) => class=fish
else => class=other

Naive Bayes

Multilayer Perceptron (ANN)
Comprehensibility - importance

**Recognized as important** since eighties [6, 7]:
- Comprehensibility is a decisive factor when ML models are applied in industry [8].
- One of the main features of ID3-like algorithms is their ability to generate easy to understand decision trees [9].

**Comprehensible classifiers enable** [10]:
- classification and classification explanation,
- classifier validation,
- knowledge discovery,
- classifier improvement and refinement.

**Domains that require comprehensible classifiers** [11]:
- credit scoring,
- medicine,
- churn prediction,
- bioinformatics.
Why classifier comprehensibility survey?

Classifier comprehensibility metric is needed to:
• compare learning systems performance and
• as a heuristic function of learning algorithm [12, 13].

Comprehensibility metric based on:
• definition (ill-defined)
• simple approximation, e.g. complexity (over-simplistic)
• expert opinion (no progress in 30 years)
• user opinion:
  • Interview (time consuming, risk of bias)
  • survey
Survey design
Survey design

6 tasks with several questions (different tree):
- objective performance measures in 4 tasks
- subjective opinion in 2 tasks

Comprehensibility parameters:
- number of leaves in a tree
- branching factor
- depth of the tree
- depth of the question/answer
- tree visualization style
- visualized/presented information
- tree layout

Tree structure

Tree visualization
## Survey design – tasks and measures

<table>
<thead>
<tr>
<th>Tasks</th>
<th>Objective performance measures</th>
<th>Subjective respondents’ opinion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classify an instance</td>
<td>Time to answer</td>
<td>Difficulty of a question</td>
</tr>
<tr>
<td>Explain classification</td>
<td>% of correct answers</td>
<td></td>
</tr>
<tr>
<td>Validate: tree confirms domain knowledge</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discover new knowledge: properties of outliers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rate tree comprehensibility</td>
<td>/</td>
<td>Comprehensibility of a tree</td>
</tr>
<tr>
<td>Compare comprehensibility of various visualizations</td>
<td>/</td>
<td>Relative comprehensibility</td>
</tr>
</tbody>
</table>

Based on [10] and [14].
Survey design

**Online survey** enables:

- accurate measurements of time (JavaScript),
- remote participation,
- translation to several languages (SVG, PHP),
- automatic checking the correctness of answers (PHP),
- saving the answers in a database (MySQL).
Survey design

Preventing bias:

- instructions and test question for each task;
- appropriate domain;
- change a single tree property;
- maximal readability;
- tasks from simple to difficult;
- subjective opinion in the last tasks;
- Latin-square design for question order;
- scales with explanations;
- statistics motivate respondents (competition).
Survey implementation
Dataset

**Zoo dataset** from UCI Machine Learning Repository [15]:
- enables learning of classifiers with various properties
- 15 binary and 1 numeric attribute
- 7 classes
- familiar to all respondents
- interesting domain for motivation
<table>
<thead>
<tr>
<th></th>
<th>antelope</th>
<th>crow</th>
<th>frog</th>
<th>honeybee</th>
<th>lobster</th>
<th>piranha</th>
<th>pitviper</th>
</tr>
</thead>
<tbody>
<tr>
<td>hair</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>feathers</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>eggs</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>milk</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>airborne</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>aquatic</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>predator</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>toothed</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>backbone</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>breathes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>venomous</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>fins</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>legs</td>
<td>4</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>6</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>tail</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>domestic</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>catsize</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>class</td>
<td>mammal</td>
<td>bird</td>
<td>amphibian</td>
<td>insect</td>
<td>mollusc</td>
<td>fish</td>
<td>reptile</td>
</tr>
</tbody>
</table>
Number of leaves in a tree
Depth of the tree

Feature 10: No backbone

Feature 11: Not airborne, doesn't breathe

Feature 8: Aquatic, doesn't breathe

Feature 4: Backbone terrestrial

Feature 9: No backbone, not airborne

Feature 5: Feathers

Feature 12: Backbone aquatic, no fins

Feature 6: Predator

Feature 7: Not predator, 6 legs

Feature 13: No backbone, airborne

Feature 14: Backbone aquatic, fins

Feature 15: Predator, not predator, 0 legs

Feature 16: Not predator, 6 legs
Tree visualization style 1/2

milk = no
|   | feathers = no
|   |   | fins = no
|   |   |   | backbone = yes
|   |   |   | aquatic = no: reptile (4.0)
|   |   |   | aquatic = yes: amphibian (5.0)
|   |   |   |   | backbone = no
|   |   |   |   | airborne = no: mollusc (12.0/2)
|   |   |   |   |   | airborne = yes: insect (6.0)
|   |   |   |   |   | fins = yes: fish (13.0)
|   |   | feathers = yes: bird (20.0)
| milk = yes: mammal (41.0)
Tree visualization style 2/2
Visualized/presented information

- milk: yes
- feathers: no
- fins: no
- backbone: no
- airborne: no
- predator: no
- insect: yes
- mollusc: yes
- aquatic: yes
- fish: yes
- bird: yes
- mammal: yes
- reptile: no
- amphibian: no
- insect: yes
- predator: no
- mollusc: yes
Tree for discover task - outliers

- milk
  - yes: mammal
  - no: airborne
    - yes: bird
    - no: legs
      - yes: fish
      - no: amphibian
Classify

Classify the instance described in the table on the left using the classification tree on the right.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>airborne</td>
<td>no</td>
</tr>
<tr>
<td>aquatic</td>
<td>no</td>
</tr>
<tr>
<td>backbone</td>
<td>yes</td>
</tr>
<tr>
<td>breathes</td>
<td>yes</td>
</tr>
<tr>
<td>feathers</td>
<td>no</td>
</tr>
<tr>
<td>fins</td>
<td>no</td>
</tr>
<tr>
<td>hair</td>
<td>yes</td>
</tr>
<tr>
<td>legs</td>
<td>4</td>
</tr>
<tr>
<td>milk</td>
<td>yes</td>
</tr>
<tr>
<td>predator</td>
<td>no</td>
</tr>
</tbody>
</table>

The instance belongs to class mammal.

Mark how difficult was the question:

- Very easy - I answered without any problems in less than 7 seconds
- Easy - I found the answer quite quickly and without major problems
- Medium -
- Difficult - I had to think hard and am not sure if I answered correctly
- Very difficult - Despite thinking very hard my answer is likely to be wrong
Statistics for motivation

**Task 1: Classify an instance**
Statistic results for the task Classify an instance:

<table>
<thead>
<tr>
<th>person</th>
<th>correct answers</th>
<th>questions answered</th>
<th>easiest question (s)</th>
<th>average time (s)</th>
<th>most difficult question (s)</th>
<th>total time (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>best (out of 61)</td>
<td>30 (100%)</td>
<td>30</td>
<td>4.4</td>
<td>11</td>
<td>19.5</td>
<td>5.5</td>
</tr>
<tr>
<td>you</td>
<td>5 (16.7%)</td>
<td>6</td>
<td>15.3</td>
<td>42.9</td>
<td>151.8</td>
<td>4.3</td>
</tr>
<tr>
<td>average (out of 61)</td>
<td>29.2 (97.3%)</td>
<td>30</td>
<td>6.9</td>
<td>19.1</td>
<td>79</td>
<td>9.6</td>
</tr>
<tr>
<td>worst (out of 61)</td>
<td>19 (63.3%)</td>
<td>30</td>
<td>14.6</td>
<td>40.4</td>
<td>680.9</td>
<td>20.2</td>
</tr>
</tbody>
</table>
Survey design validation
Survey Design Validation

- The time to answer a question is correlated with the rated difficulty.
- The % of correct answers is correlated with difficulty in two tasks.
- Survey should be more understandable to non-experts in DM.
- The 1st task (classify) is the easiest, the 4th (discover) the hardest.
- The range of questions according to difficulty in each task is broad.

<table>
<thead>
<tr>
<th>task</th>
<th>time-difficulty correlation</th>
<th>correct-difficulty correlation</th>
<th>correct answers (%)</th>
<th>question time (s)</th>
<th>question difficulty</th>
</tr>
</thead>
<tbody>
<tr>
<td>classify</td>
<td>0.772</td>
<td>0.094</td>
<td>98.1</td>
<td>8.6 16.7 31.4</td>
<td>1.33 1.63 2.06</td>
</tr>
<tr>
<td>explain</td>
<td>0.957</td>
<td>-0.432</td>
<td>92.0</td>
<td>8.4 24.7 50.8</td>
<td>1.50 2.02 2.61</td>
</tr>
<tr>
<td>verify</td>
<td>0.720</td>
<td>-0.658</td>
<td>96.4</td>
<td>7.6 14.9 22.1</td>
<td>1.50 1.95 2.33</td>
</tr>
<tr>
<td>discover</td>
<td>0.901</td>
<td>0.548</td>
<td>62.5</td>
<td>12.7 28.6 44.6</td>
<td>2.00 2.53 2.82</td>
</tr>
</tbody>
</table>
Survey Design Validation – tree parameters

• The **number of leaves** is correlated with:
  • rated tree comprehensibility (0.951),
  • rated question difficulty (0.452 to 0.802),
  • time to correctly answer a question (0.418 to 0.663),
  • tree with 4 leaves rated more comprehensible than tree with 10 leaves (2.89 on 1-4 scale).

• Similar conclusion for **depth of the tree** (lower correlations).

• Correlations with **branching factor** are close to zero:
  • the difference between branching factor 2 and 4 is not big enough or
  • increasing the branching factor does not decrease the comprehensibility.

• **Depth of question** correlated to respondents’ performance (0.66 to 0.943):
  • related work considers only comprehensibility of the whole classification tree;
  • emphasize the more important parts of the tree instead.

• **Tree presentation** influences comprehensibility:
  • **simple text** representation much more difficult to comprehend (3.72 on 1-4 scale),
  • **meaningful names** of classes and attributes are more comprehensible (rate 2.44),
  • **tree layout** and **Weka vs. Orange** output slightly influences compr. (1.89 and 1.83).
Survey Design Validation – range of trees

Good range of trees:
• comprehensibility from 1.44 to 3.83.

Appropriate scale:
• respondent’s ratings agree well.

<table>
<thead>
<tr>
<th>Tree Properties</th>
<th>Comprehensibility</th>
</tr>
</thead>
<tbody>
<tr>
<td>L=3 B=2 D=2</td>
<td></td>
</tr>
<tr>
<td>L=3 B=3 D=1</td>
<td></td>
</tr>
<tr>
<td>L=4 B=2 D=3</td>
<td></td>
</tr>
<tr>
<td>L=4 B=4 D=1</td>
<td></td>
</tr>
<tr>
<td>L=7 B=2 D=5</td>
<td></td>
</tr>
<tr>
<td>L=7 B=3 D=3</td>
<td></td>
</tr>
<tr>
<td>L=7 B=4 D=2</td>
<td></td>
</tr>
<tr>
<td>L=9 B=3 D=2</td>
<td></td>
</tr>
<tr>
<td>L=9 B=3 D=4</td>
<td></td>
</tr>
<tr>
<td>L=10 B=2 D=7</td>
<td></td>
</tr>
<tr>
<td>L=10 B=4 D=3</td>
<td></td>
</tr>
</tbody>
</table>

Legend:
- very easy (1)
- easy (2)
- medium (3)
- difficult (4)
- very difficult (5)
Comprehensibility of Classification Trees – Survey Design Validation

rok.piltaver@ijs.si

Rok Piltaver, Mitja Luštrek, Matjaž Gams
Jozef Stefan Institute - Department of Intelligent Systems
Jozef Stefan International Postgraduate School

Sanda Martinčić - Ipšić
University of Rijeka - Department of Informatics

6th International Conference on Information Technologies and Information Society
Šmarješke toplice, Slovenia, November 6-7, 2014
References


[9] D. Michie, EWSL'87 in Bled


