### Visual Data Mining

<table>
<thead>
<tr>
<th>Feature</th>
<th>Data Mining Algorithms</th>
<th>Visualization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actionable</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Evaluation</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Flexibility</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>User Interaction</td>
<td>-</td>
<td>+</td>
</tr>
</tbody>
</table>

**Diagram:**
- **Data Mining**
- **Visual Data Mining**
- **Information Visualization**
Visual Data Mining Architecture:
Tightly Integrated Visualization
One example for Preceding Visualization

Circle Segments Visualization of Stock Data

- Exploring ~10,000 records
- 50 different stock prices
Two examples for Subsequent Visualization

- SPSS AnswerTree
- Decision Tree Visualizer (MineSet)
Tightly Integrated Visualization

Visualization of algorithmic decisions

- Data and patterns are better understood
- User can make decisions based on perception
- User can make decisions based on domain knowledge
- Visualization of result enables user specified feedback for next algorithmic run
Tightly Integrated Visualization

• The first prototypes which follow this architecture:

  - Perception-Based Classification (Decision Tree Classification)
    • HD-Eye (Clustering)
    • DataJewel (Temporal Mining)
The corresponding DM Method

Classification

Problem description:
Given a set of objects with known class labels.

➢ Description
Build model describing the data with respect to the class

➢ Prediction
Use model to predict the class label of objects
**Problem description:**
Given data describing individuals, find factors that indicate buyers.

**Algorithm**
- Search through all factors to find one which best divides people into buyers / not buyers
- Divide groups and repeat on subgroups

**Outcome**
Tree uses factors to describe people who are likely to be buyers
Tutorial (2)
Extracting Rules from a Decision Tree

IF (age < 35) THEN Class = ‘Yes’

IF (age >= 35) AND (sal < 67) THEN Class = ‘No’

IF (age >= 35) AND (sal >= 67) THEN Class = ‘Yes’
### Visual Classification

- Each attribute is sorted and visualized separately.
- Each attribute value is mapped onto a unique pixel.
- The color of a pixel is determined by the class label of the object.
- The order is reflected by the arrangement of the pixels.

<table>
<thead>
<tr>
<th>attr.1</th>
<th>attr.2</th>
<th>...</th>
<th>class</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.3</td>
<td>23.3</td>
<td>...</td>
<td>Y</td>
</tr>
<tr>
<td>2.4</td>
<td>2.0</td>
<td>...</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>attr. 1</th>
<th>class</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>Y</td>
</tr>
<tr>
<td>0.3</td>
<td>Y</td>
</tr>
<tr>
<td>0.3</td>
<td>Y</td>
</tr>
<tr>
<td>0.5</td>
<td>N</td>
</tr>
<tr>
<td>1.1</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>attr. 2</th>
<th>class</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>N</td>
</tr>
<tr>
<td>1.3</td>
<td>Y</td>
</tr>
<tr>
<td>2.0</td>
<td>N</td>
</tr>
<tr>
<td>2.5</td>
<td>Y</td>
</tr>
<tr>
<td>5.1</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Visual Classification

Attribute 1
Attribute 2

Pixel Arrangement
PBC Color Scale

An Example:

n-ary split
binary split
Visual Classification

- Shuttle data set (9 attributes, 43,500 records)
Visual Classification

- Segment data set (19 attributes, 7 classes)
Visual Classification

A New Visualization of a Decision Tree

age < 35
Visual Classification

A New Visualization of a Decision Tree

age < 35

G

Salary

< 40 [40,80] > 80
Visual Classification

A New Visualization of a Decision Tree

age < 35

< 40
[40,80]
> 80

G
V
P

Salary
Visual Classification

A Decision Tree for the Segment Data Set
Visual Classification

Different types of algorithmic support for the user:

- Propose split
- Look-ahead
- Expand subtree
Visual Classification

- Animated split lines
- Magnified split lines
- Exact split point
## Visual Classification

### Accuracy:

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Australian</td>
<td>84.9</td>
<td>80.9</td>
<td><strong>86.9</strong></td>
<td>82.7</td>
</tr>
<tr>
<td>DNA</td>
<td><strong>93.8</strong></td>
<td>89.2</td>
<td>93.3</td>
<td>89.2</td>
</tr>
<tr>
<td>Satimage</td>
<td>86.4</td>
<td>84.1</td>
<td><strong>86.8</strong></td>
<td>83.5</td>
</tr>
<tr>
<td>Segment</td>
<td>95.5</td>
<td>95.0</td>
<td><strong>96.3</strong></td>
<td>94.8</td>
</tr>
<tr>
<td>Shuttle</td>
<td>99.5</td>
<td>99.6</td>
<td>99.7</td>
<td><strong>99.9</strong></td>
</tr>
</tbody>
</table>
Conclusions

- Tight integration of visualization and data mining algorithms is still a very new area of research

- Data mining algorithms and visualization technique can nicely complement each other.

- PBC leverages decision tree algorithms, allows the user to steer the mining process.

- User involvement during the mining process enables knowledge transfer and capitalizes on human’s perception