Visual Data Mining

Apr. 24, 2007

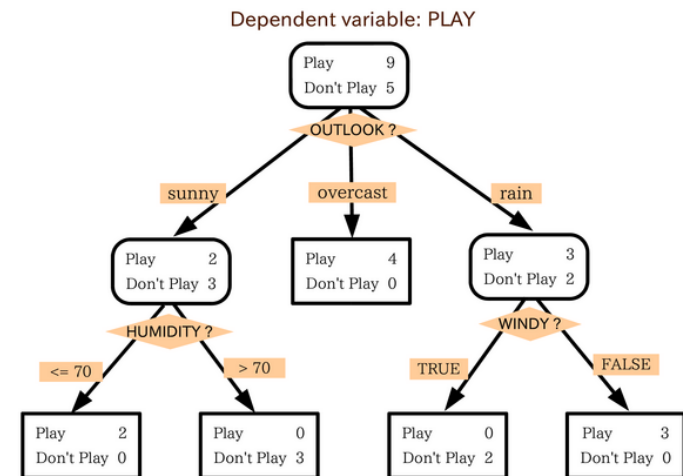Visual Analytics presentation

Julia Nam

Overview

- **Visual Classification: An Interactive Approach to Decision Tree Construction**
  M. Ankerst, C. Elsen, M. Ester, H. Kriegel, University of Munich

- **StarClass: Interactive Visual Classification Using Star Coordinates**
  Soon Tee Teoh, Kwan-Liu Ma, University of California, Davis

- **PaintClass: Interactive Construction, Visualization and Exploration of Decision Trees**
  Soon Tee Teoh, Kwan-Liu Ma, University of California, Davis

---

Decision Trees

Play golf dataset

OUTLOOK	TEMPERATURE	HUMIDITY	WINDY	PLAY
sunny	85	85	FALSE	Don't Play
sunny	80	90	TRUE	Don't Play
overcast	83	78	FALSE	Play
rain	70	96	FALSE	Play
rain	68	80	FALSE	Play
rain	65	70	TRUE	Don't Play
overcast	64	65	TRUE	Play
sunny	72	95	FALSE	Don't Play
sunny	69	70	FALSE	Play
rain	75	80	FALSE	Play
sunny	75	70	TRUE	Play
overcast	72	90	TRUE	Play
overcast	81	75	FALSE	Play
rain	71	80	TRUE	Don't Play

Independent variables / Dep. var

## Decision Trees

Dependent variable: PLAY

Decision Tree Classifier

- Most classification systems are designed for minimal user intervention.

- Using visualization, they allows to integrate the domain knowledge of an expert in the tree construction phase.

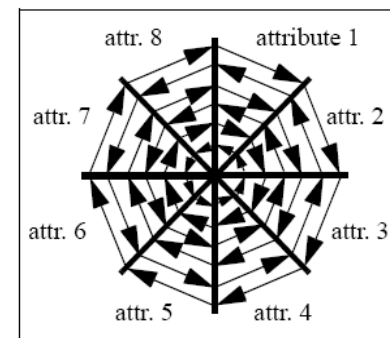- Also, they allow to explore the decision tree interactively.

Visual Classification:
An Interactive Approach to Decision Tree Construction
(KDD '99)

M. Ankerst, C. Elsen, M. Ester, H. Kriegel,
University of Munich
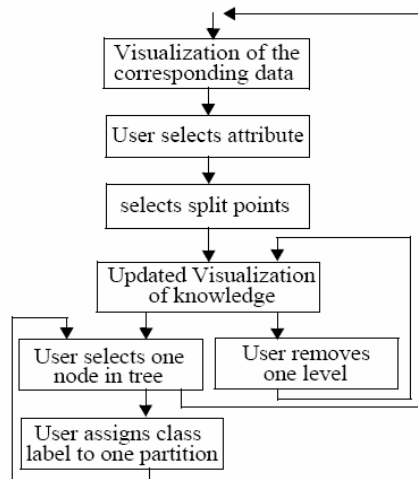
PBC (Perception Based Classification)
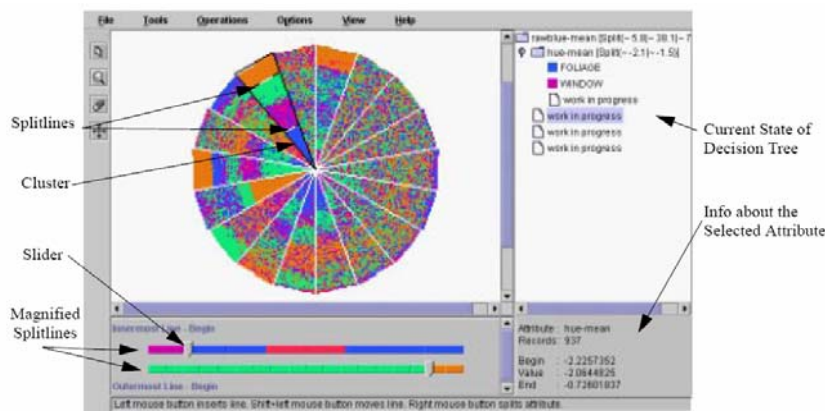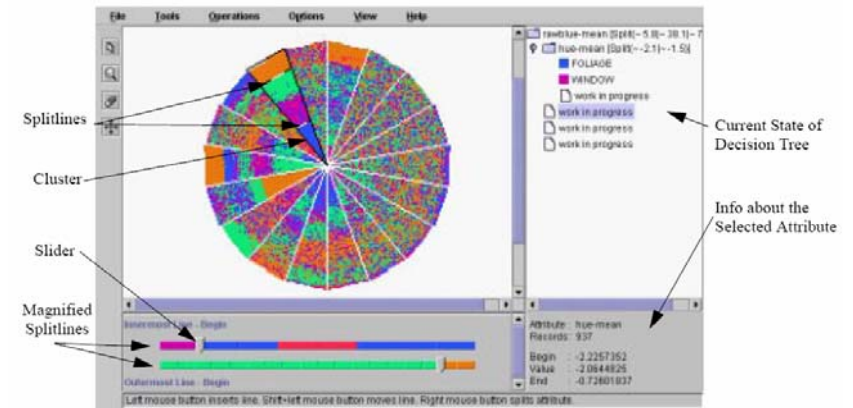

Circle Segments

- They represent all values of one attribute in a segment of a circle with the proposed arrangement inside a segment.

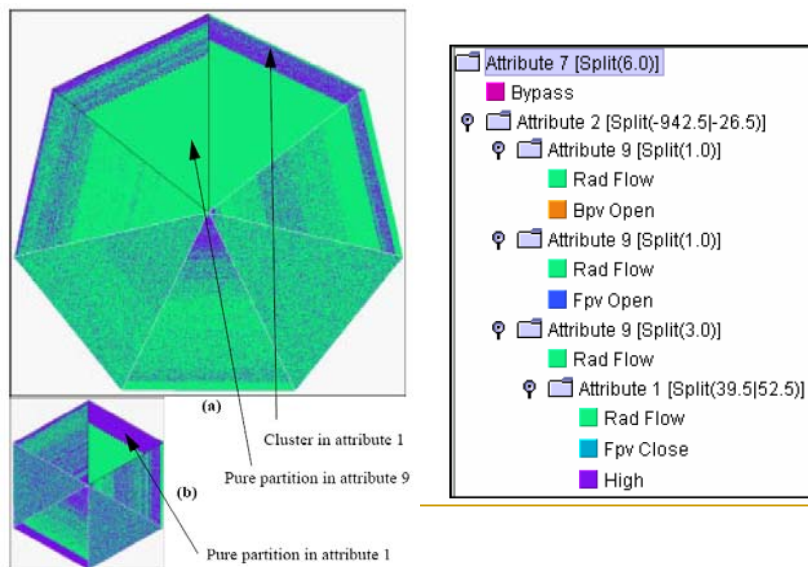- The color of pixel is determined by the class label of the object.

Interactive Classification

PBC (Perception Based Classification)

Experimental Evaluation

- Cluster in attribute 1
- Pure partition in attribute 9
- Pure partition in attribute 1

Experimental Evaluation

Accuracy	CART	C4	SPRINT	CLOUDS	PBC
Satimage	85.3	85.2	86.3	85.9	83.5
Segment	94.9	95.9	94.6	94.7	94.8
Shuttle	99.9	99.9	99.9	99.9	99.9

Tree size	CART	C4	SPRINT	CLOUDS	PBC
Satimage	90	563	159	135	60
Segment	52	102	18.6	55.2	39.5
Shuttle	27	57	29	41	14.6

Table1,Table 2: Accuracy and Tree size of PBC and algorithmic approaches

Contribution

- Introduce a fully interactive method for decision tree construction based on a multidimensional visualization.

- Domain knowledge of an expert can be profitably included in the tree construction phase.

- After that, the user has a much deeper understanding of the data than just knowing the decision tree generated by an arbitrary algorithm.

StarClass & PaintingClass: Interactive Construction, Visualization and Exploration of Decision Trees
(Conf. On Data Mining SDM '03 & SIGKDD '03)

Soon Tee Teoh, Kwan-Liu Ma
*University of California, Davis*

PaintingClass

- Start with a set of training data.

- Every object in the set is projected and displayed visually into 2-D space.

- The user creates a projection that best separates the data objects belonging in different classes.

- Then it is partitioned by the user into regions.

- Repeat projection and partitioning until the user has a satisfactory decision tree.

- The class is leaf nodes of the tree.

Visual Projections in PaintingClass

- Star Coordinates
  - Good at showing dimensions with numerical attributes
  - StarClass: Interactive Visual Classification Using Start Coordinates

- Parallel Coordinates
  - Good at showing dimensions with categorical attributes

Star Coordinates
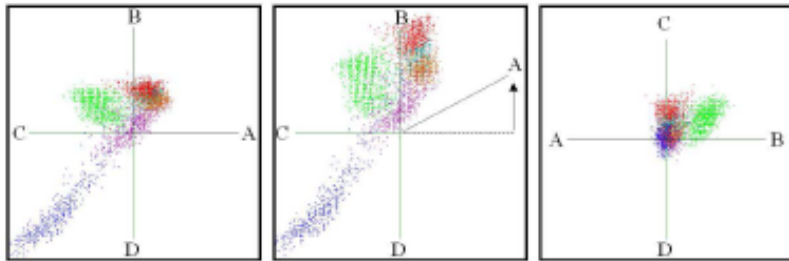
- User can select any projection.
- The color of the data object is the assigned class.

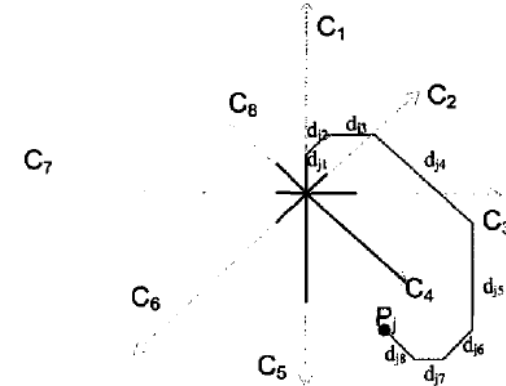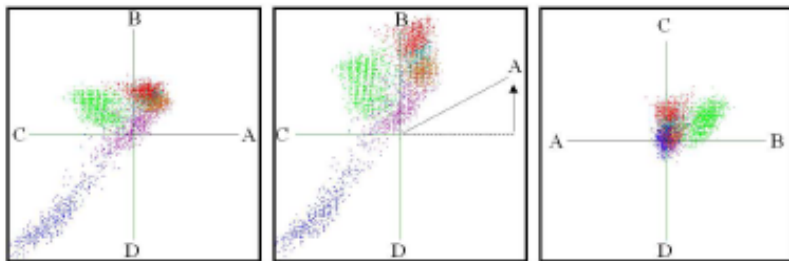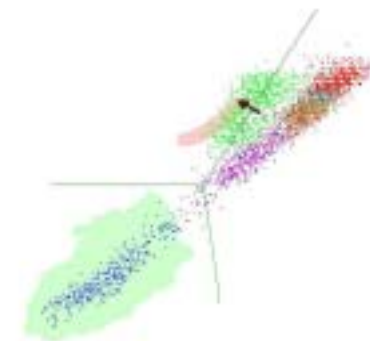$$D_j = \left(d_{j0}, d_{j1}, \ldots, d_{jl}, \ldots, d_{jn} \right);$$

Figure 2. The location of an 8-dimensional point is the vector sum of its value on each dimension.
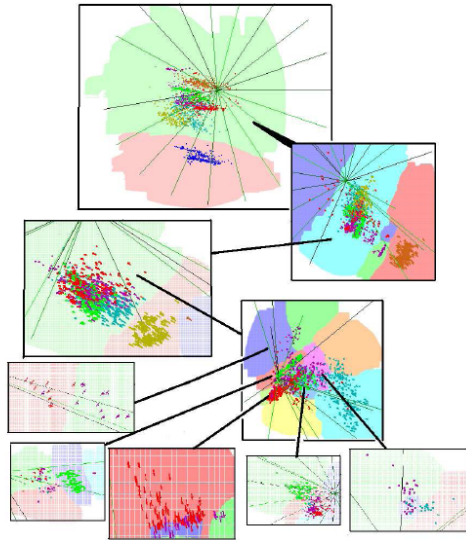
Star Coordinates

- Can scale or rotate an axis by moving an axis.
- Can generate different axis mapping.
- Can be zoomed by selecting region.
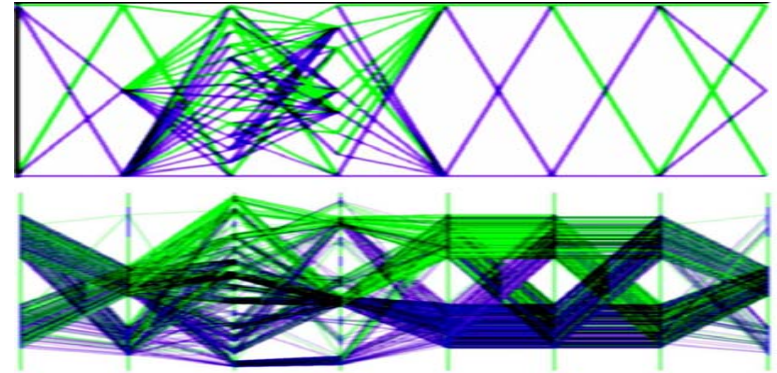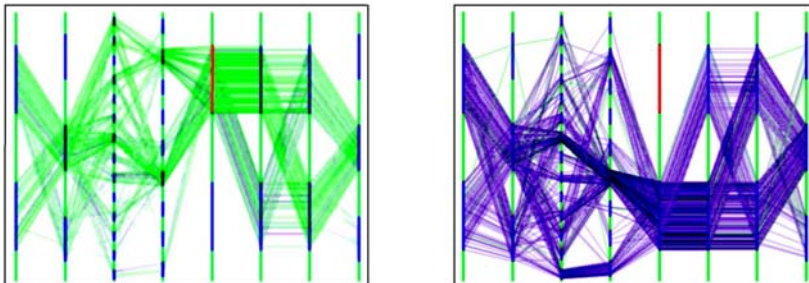
Painting Regions

Building the decision Tree
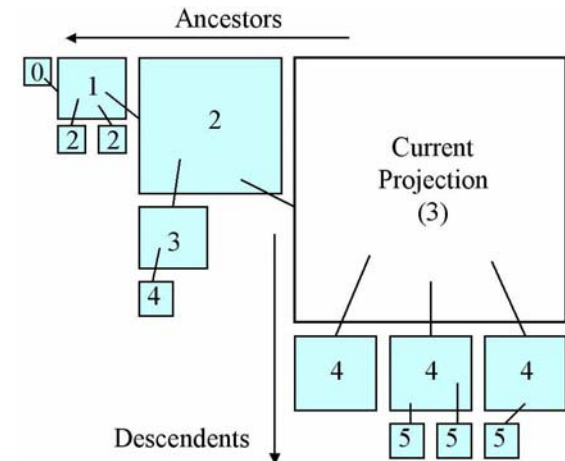
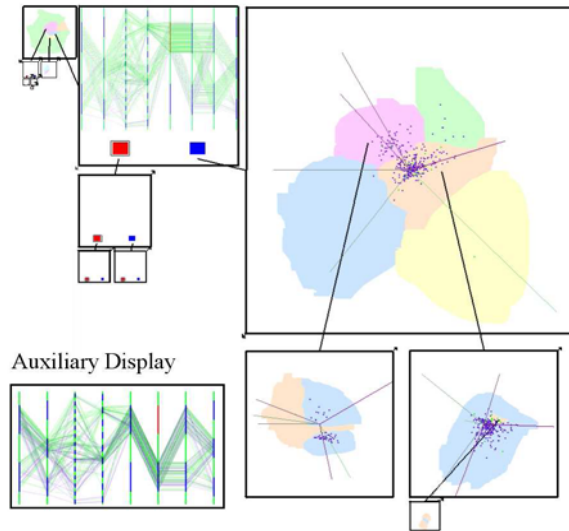

Parallel Coordinates

- To display the categorical dimensions,

Painting Regions

Decision Tree Visualization & Exploration – Focus + Context

Decision Tree Visualization & Exploration

Auxiliary Display

The Objectives of PaintingClass

- To create a user-directed decision tree classification system

- To enable users to explore and visualize multi-dimensional data and their corresponding decision trees to improve user understanding and to gain knowledge.

First Goal : Good enough Accuracy

Table 2: Accuracy of PaintingClass compared with algorithmic approaches and visual approach PBC.

	Algorithmic			Visual	
	CART	C4	SLIQ	PBC	PaintingClass
Satimage	85.3	85.2	86.3	83.5	85.3
Segment	84.9	95.9	94.6	94.8	95.2
Shuttle	99.9	99.9	99.9	99.9	99.9
Australian	85.3	84.4	84.9	82.7	84.7

Table 3: Accuracy of PaintingClass compared with other classification methods.

	CBA	C4.5	FID	Fuzzy	PaintingClass
Australian	85.0	82.6	58.0	88.9	84.7
adult	84.2	85.4	23.6	85.9	85.1
diabetes	74.4	73.8	62.0	77.6	74.6

Second Goal : Knowledge Gained

- User can see the hierarchy of split points.
  - They reveal which dimensions, combinations of dimensions and values are most correlated to different classes.
- The shape of the decision tree indicates certain characteristics of the data.
- Each node in the tree is itself a visual projection of a subset of the data.
  - It can reveal patterns, clusters, shapes and outliers.
- Hierarchy allows the user to focus on subsets of the data.

Conclusion

- The user interactively edits projections of multi-dimensional data and paints regions to build a decision tree.

- Exploration of decision trees fulfills some general goals of data mining beyond classification.

- PaintingClass also provides several useful extensions to StarClass, the most important of which is the use of Parallel Coordinates to display categorical values so that even datasets with categorical dimensions can be classified and visualized.