

Lab 0

High-Performance Data Mining: Problems and Current Tools

Monday, May 15, 2000

William H. Hsu

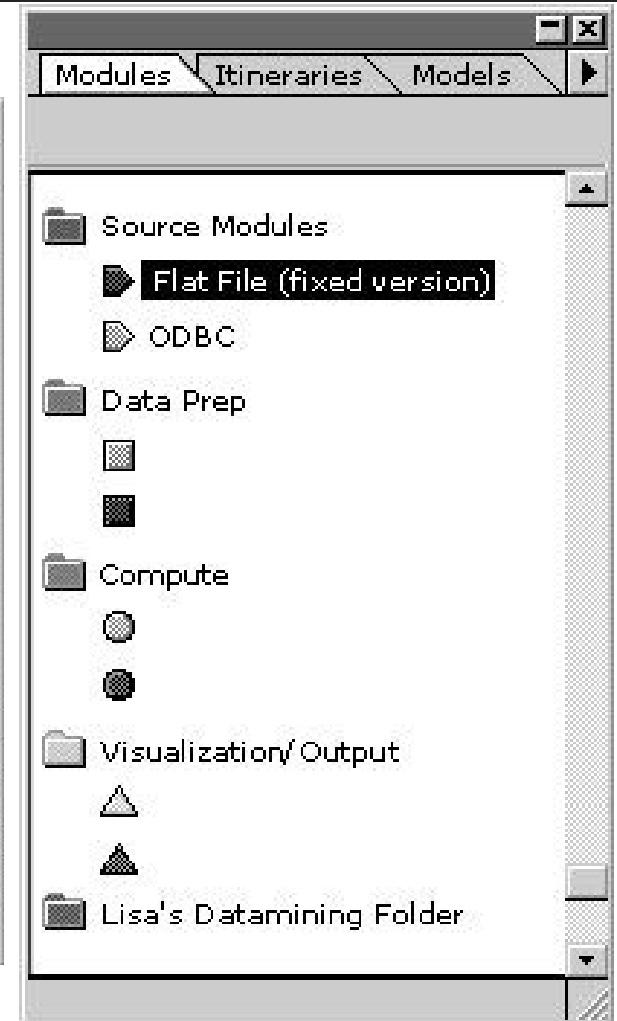
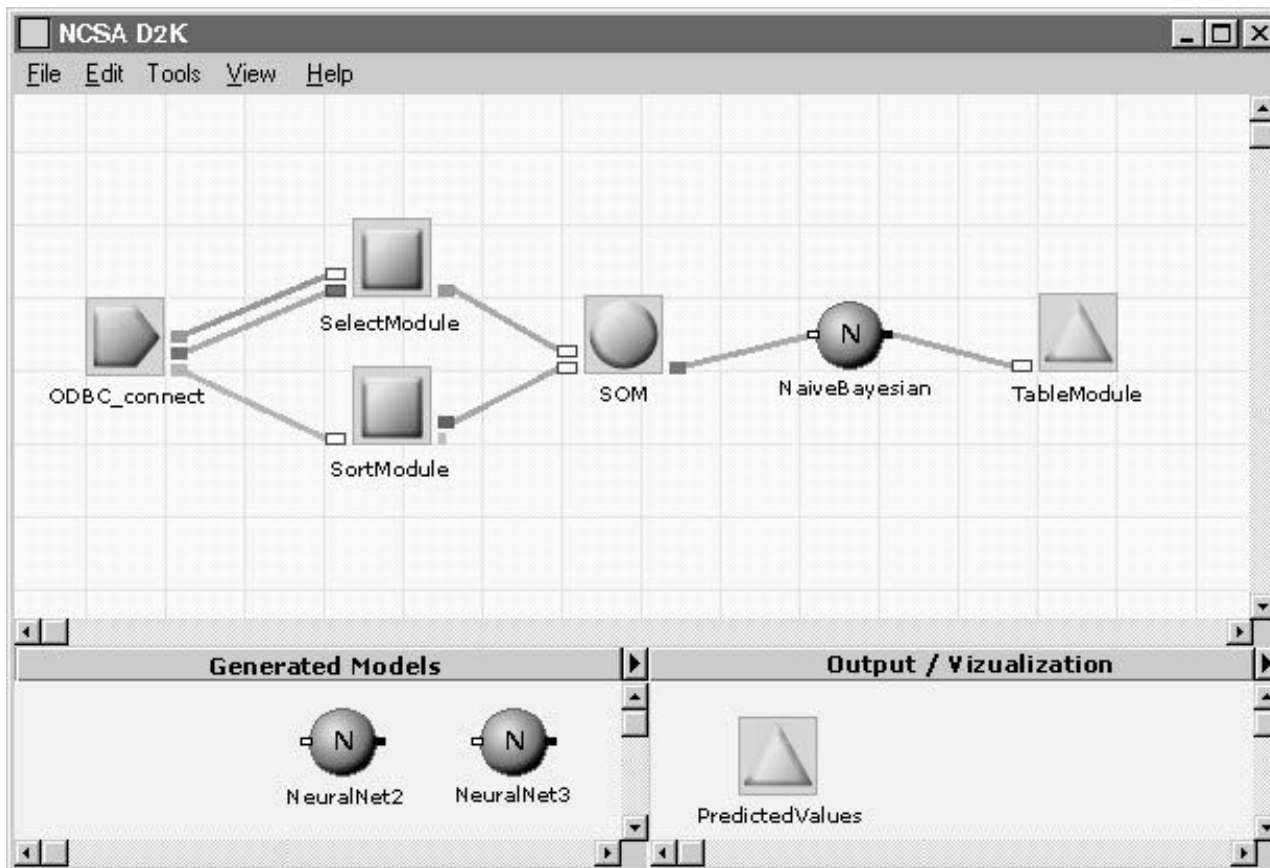
Department of Computing and Information Sciences, KSU

<http://www.cis.ksu.edu/~bhsu>

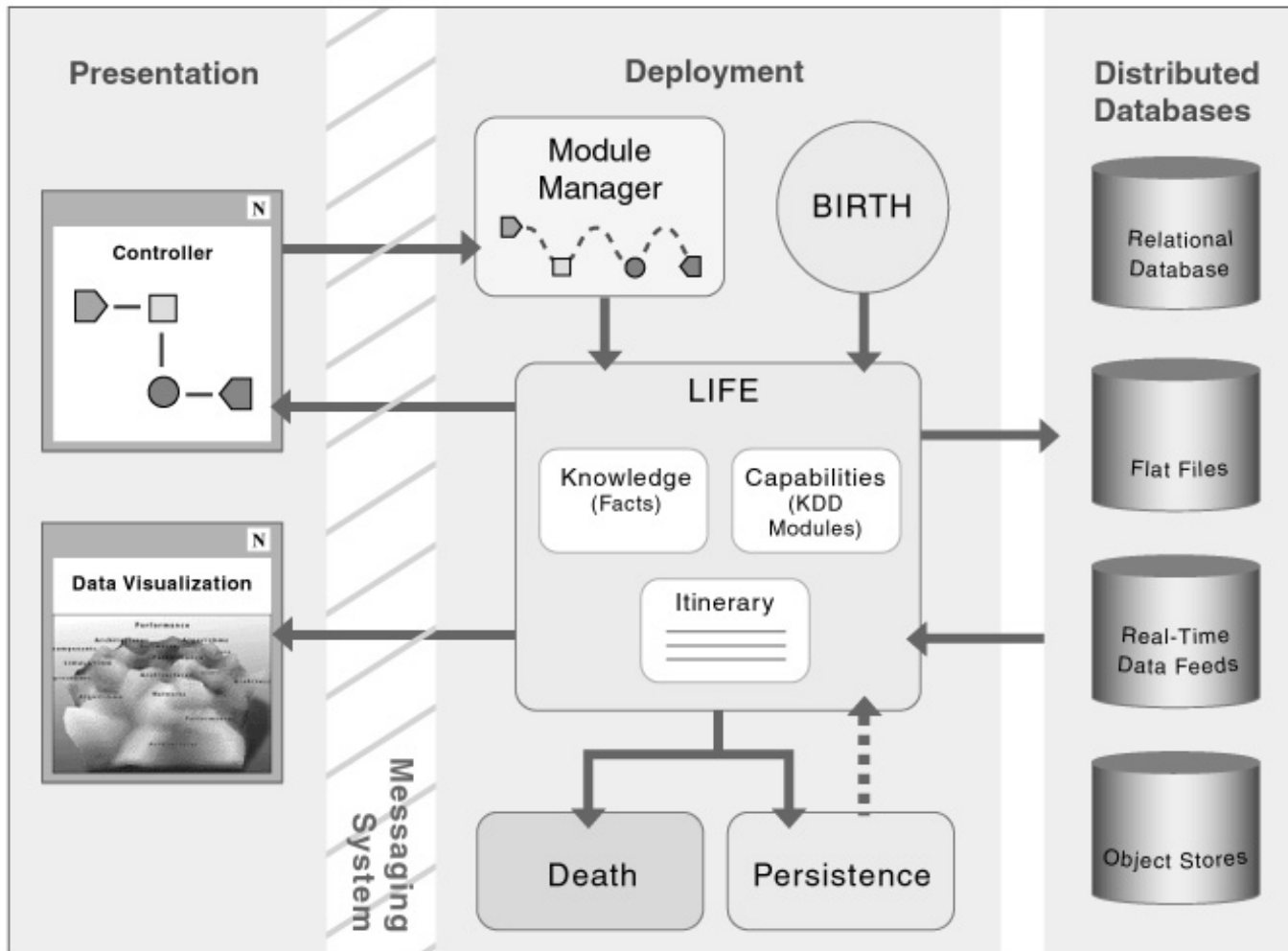
Data Mining Software Practicum

- **Lab and Implementation Project Objectives**
 - Learn to work with some KDD software tools
 - Understand modern ML-based object oriented HP-KDD systems
 - Main tools: *MLC++/MineSet* and *NCSA D2K*
- **Implementation Project Milestones**
 - Project plan
 - Due Monday, May 22, 2000 (Homework 1)
 - Short writeup: 1-2 pages design plus simple specification document
 - Project interim report
 - Due Wednesday, May 31, 2000 (Homework 2)
 - Preliminary itinerary and documentation
 - Comparative results
 - Using software tools
 - Due Thursday, June 8, 2000 (machine problem; Homework 3)
 - Exposure to: *MineSet*; *SNNS*, *NS3*; *Hugin*, *BKD*; *GPSys*, *Genesis*

Software Environments for KDD: *NCSA D2K*



KDD and Software Engineering: D2K Framework



Rapid KDD Development Environment

Performance Element: Decision Support Systems (DSS)

- **Model Identification (Relational Database)**

- Specify data model
- Group attributes by type (dimension)
- Define queries

- **Prediction Objective Identification**

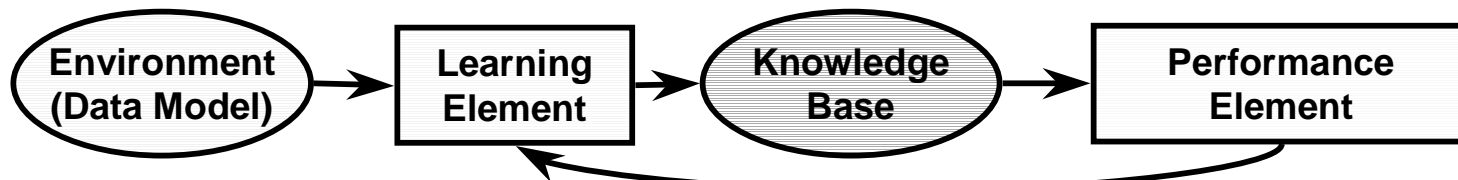
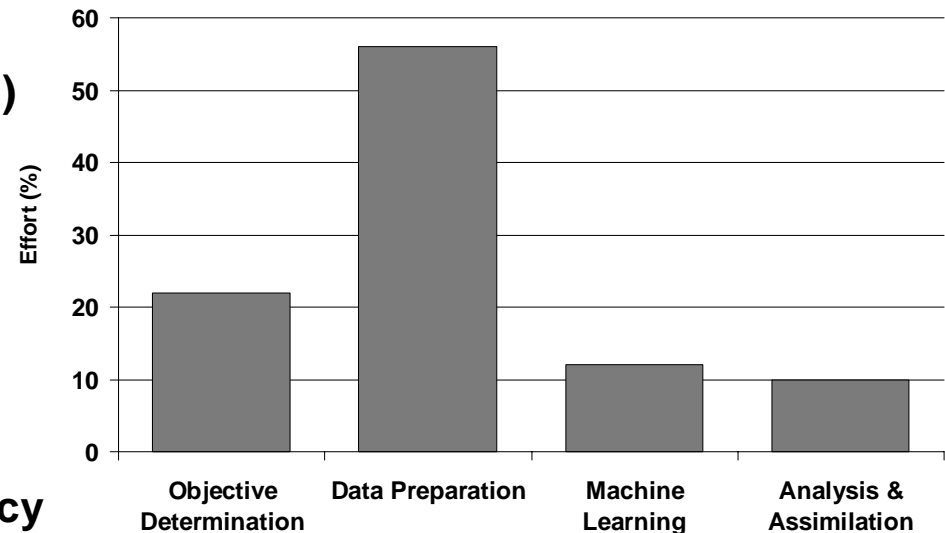
- Identify target function
- Define hypothesis space

- **Transformation of Data**

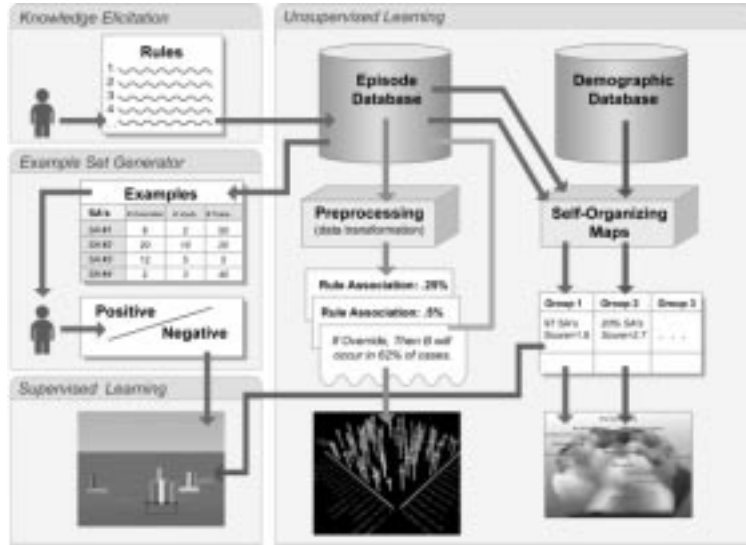
- Reduce data: e.g., decrease frequency
- Select *relevant* data channels (given prediction objective)
- Integrate models, sources of data (e.g., *interactively elicited rules*)

- **Supervised Learning**

- **Analysis and Assimilation: Performance Evaluation using DSS**

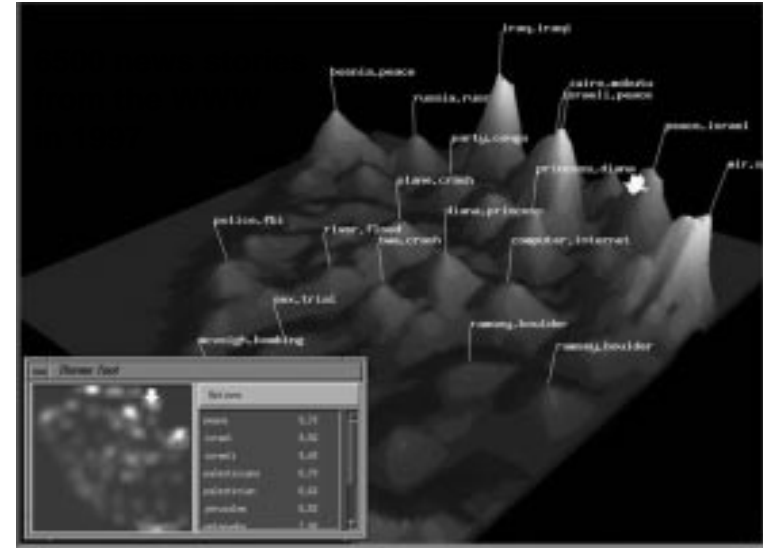


Some Interesting Industrial Applications



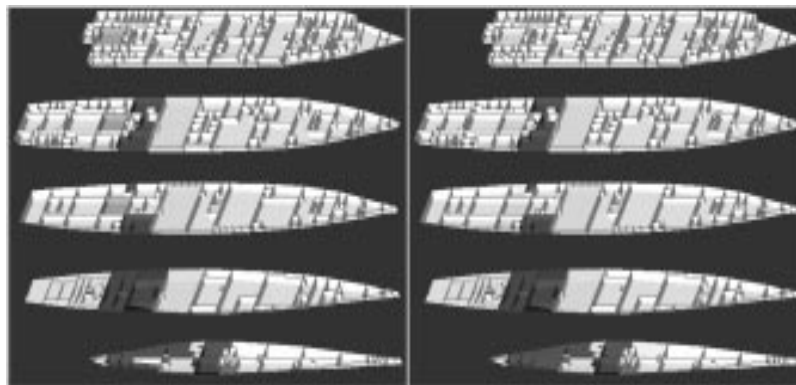
NCSA D2K - <http://www.ncsa.uiuc.edu/STI/ALG>

Database Mining



Cartia ThemeScapes - <http://www.cartia.com>

Reasoning (Inference, Decision Support)



DC-ARM - <http://www-kbs.ai.uiuc.edu>

Normal

Ignited

Engulfed

Extinguished

Fire Alarm

Floding

Planning, Control

