Comprehensible Data Mining: Gaining Insight from Data

Michael J. Pazzani
Information and Computer Science
University of California, Irvine
pazzani@ics.uci.edu
http://www.ics.uci.edu/~pazzani

Outline

- **UC Irvine's data mining program**
- **KDD:**
  - **Goals: Gaining insight from data**
  - **Methods: Learn predictive and/or descriptive models**
  - **Conclusion: Not all models provide "insight"**
    - » **Validate Findings**
    - » **Deliver Findings**
- **Comprehensibility and Prior Knowledge**
  - **Expert IF/Then Rules**
  - **Monotonocity constraints**
  - **Negative Interactions**
- ***Knowledge placed in the perspective of what is already known. - Dr Ruth David***

University of California, Irvine

- Ph.D and M.S. NEW with focus on data mining
  - Rina Dechter   Bayesian Networks
  - Richard Granger   Neural Networks
  - Dennis Kibler   Inductive Learning
  - Richard Lathrop   Learning and Molecular Biology
  - Michael Pazzani   Knowledge-intensive learning
  - Padhraic Smyth   Probabilistic Models & KDD
- Archive of over 100 databases used in learning research http://www.ics.uci.edu/~mlearn
- "Proprietary" databases analyzed in conjunction with sponsors

Applications

- Telephone(NYNEX)- Diagnosis of local loop.
- Economic Sanctions (RAND)- Predict whether economic sanctions will have desired goal.
- Foreign Trade Negotiations (ORD)- Predict conditions under partner will make a concession.
- Pharmaceutical-
- Dementia- (UCI and CERAD)- Screening for Alzheimer’s disease. Cognitive and Functional questionnaires
- Supermarket scanner data
- User Profiles- text & demographics

Summary

- A variety of techniques can learn predictive models that exceed or rival the performance of human experts
- Demonstrating predictive accuracy is not sufficient for adopting a predictive model.
- Experts will not gain any insight from a relationship that they don't believe
- Signs of acceptance
  - Publication in peer-reviewed journals
  - Adopted in practice
- Experts give more credence to models that don't unnecessarily violate prior expectations

Economic Sanctions

- *In 1983, Australia refused to sell uranium to France, unless France ceased nuclear testing in the South Pacific. France paid a higher price to buy uranium from South Africa.*

- *In 1980, the US refused to sell grain to the Soviet Union unless the Soviet Union withdrew troops from Afghanistan. The Soviet Union paid a higher price to buy grain from Argentina and did not withdraw from Afghanistan.*

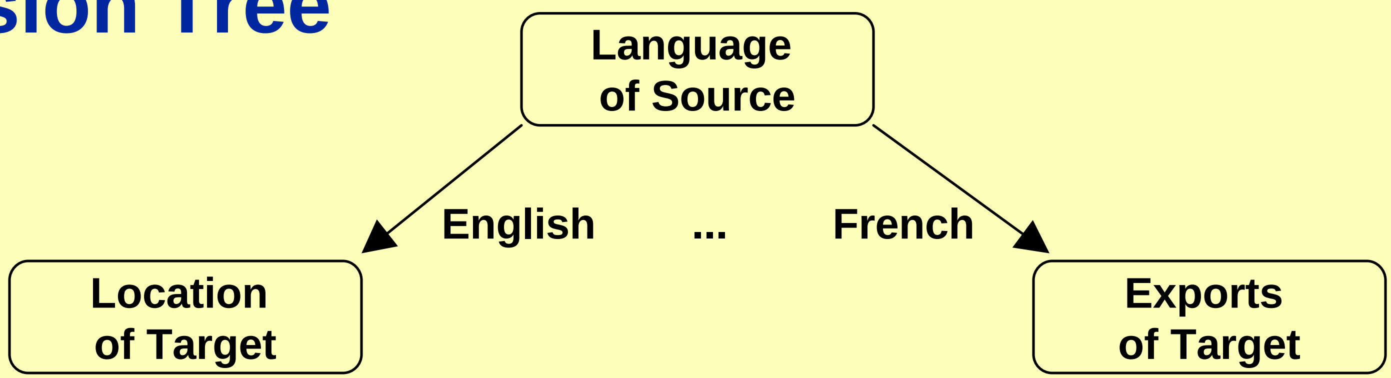
Regression

- **Predicting amount of effect of sanctions as a linear combination of variables.**
- **Hufbauer, Schott & Elliot (1985). *Economic sanctions Reconsidered.* Institute for International Economics**
- **Effect= 12.23 - 0.94SCOST + 0.17TCOST +10.26WW-0.16Cooperation-0.24 Years**

  R^2 = .21
- **Selecting and Inventing relevant variables**
- **Equation doesn't always make sense**

Learning Rules and Trees

- **Least General Generalization:**
  - If an English speaking democracy that imports oil threatens a country in the Northern Hemisphere that has a strong economic health and exports weapons, then the sanction will fail because a country in the Southern Hemisphere will sell them the product.
- **Decision Tree**



Dementia Screening

- Analysis of data collected by the Consortium to Establish a Registry for Alzheimer's Disease (CERAD)
- Distinguish "normal" or "mildly impaired" patients
- Demographic data (age, gender, education, occupation)
- Answers to Cognitive Questionnaires
  - Mini-Mental Status Exam
  - Blessed Orientation, Memory and Concentration
  - e.g., remember address: John Brown, 42 Market Street, Chicago
- Current usage is a simple threshold on the number of errors
  - If there are more than 9 mistakes, then the patient is impaired
  - Accuracy 49.0%;sensitivity 13.7%; specificity 99.27%

Learning Rules for Dementia Screening

IF the years of education of the patient is > 5
AND the patient does not know the date
AND the patient does not know the name of a nearby street
THEN The patient is NORMAL

OTHERWISE IF the number of repetitions before correctly reciting the address is > 2
AND the age of the patient is > 86
THEN The patient is NORMAL

OTHERWISE IF the years of education of the patient is > 9
AND the mistakes recalling the address is < 2
THEN The patient is NORMAL

OTHERWISE The patient is IMPAIRED

Accuracy of Learned Models

Algorithm	Accuracy
General Practitioner	~60%
Neurologists	~85%
C4.5	86.7
C4.5 rules	82.6
Naïve Bayes	88.7
FOCL	90.6

- **Although accuracy is acceptable, experts were hesitant to accept rules because they violated the intended use of the tests**
  - **Getting a question right used as sign of dementia**
  - **Getting questions wrong used as evidence against dementia.**
  - **2.13 violations for an average rule**

Comprehensibility of Learned Models

- **Pruning- Simplicity bias**
  - Delete unnecessarily complex structures
- **Visualization**
  - Interactive Exploration of Complex Structures
- **Iteration-**
  - Delete, invent variables
  - Change parameters, learning algorithm
- **Consistency with existing knowledge**
  - Strong Domain Theories
  - Weak Domain Theories
  - Association Rules

Simpler isn't always better

- Most work in ML and KDD equates "understandable" with "concise"

A. **If** the native language of the country is English
   Then the sales of leisure products will be high

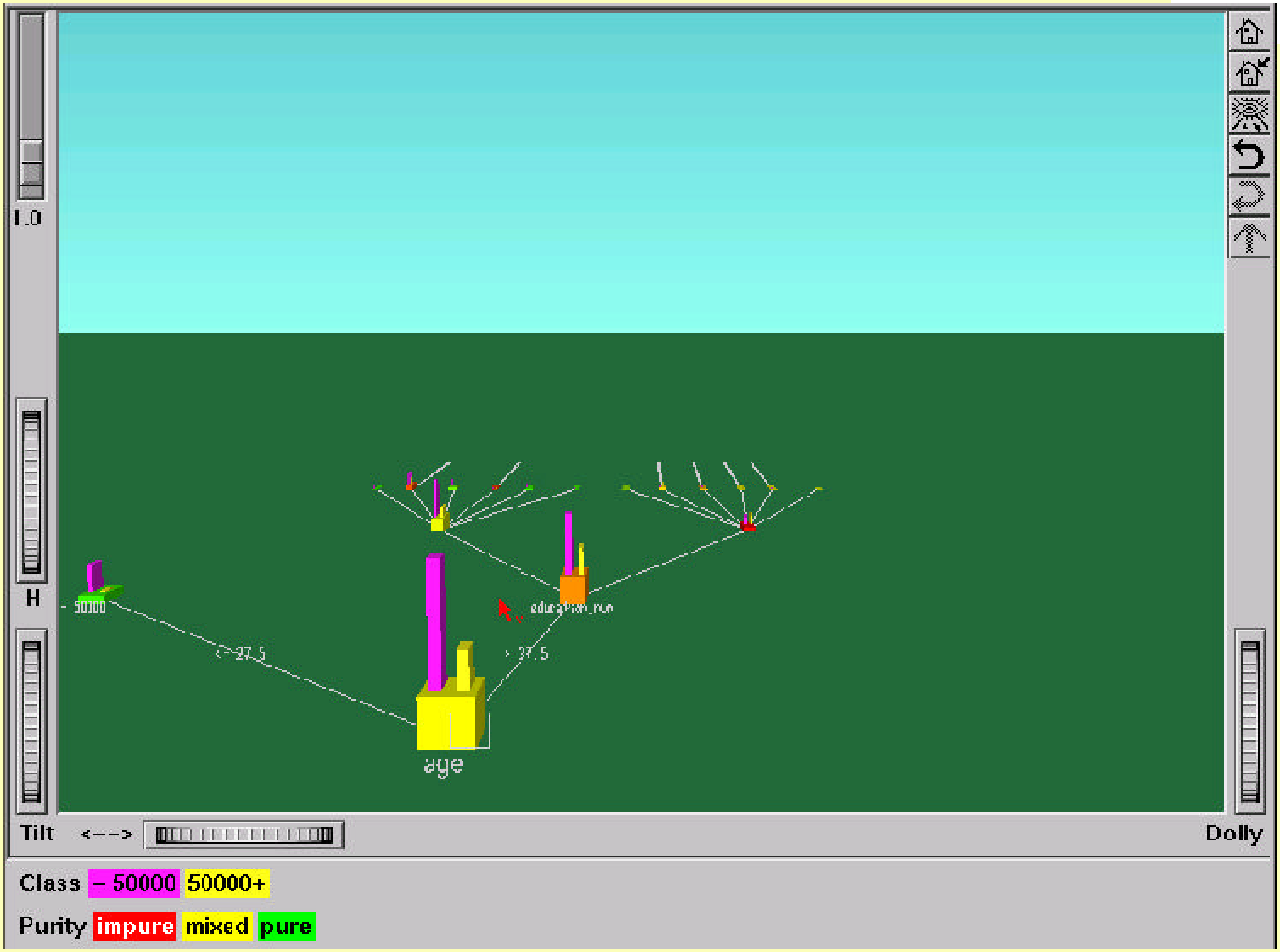
B. **If** there is a large population with high income
   and there is a free market economy
   Then the sales of leisure products will be high

- Problem- There are often many models with similar complexity consistent with the data

A. **If** the average height < 6foot6inch
   Then the the team will score on fast breaks

B. **If** the average time at 40m is < 4.2 sec
   Then the the team will score on fast breaks

Visualizing Incomprehensible Decision Trees



Comprehensibility and Prior Knowledge

- When creating models from data, there are many possible models with equivalent predictive power.
- Understandability by users should be used to constrain model selection.
- One factor that influences understandability is consistency with domain knowledge.

Explanation-based Learning:
Using Strong Domain Knowledge

- Explain why an item belongs to a class
- Retain features of examples used in explanation

If the supply of an object decreases
Then the price will increase

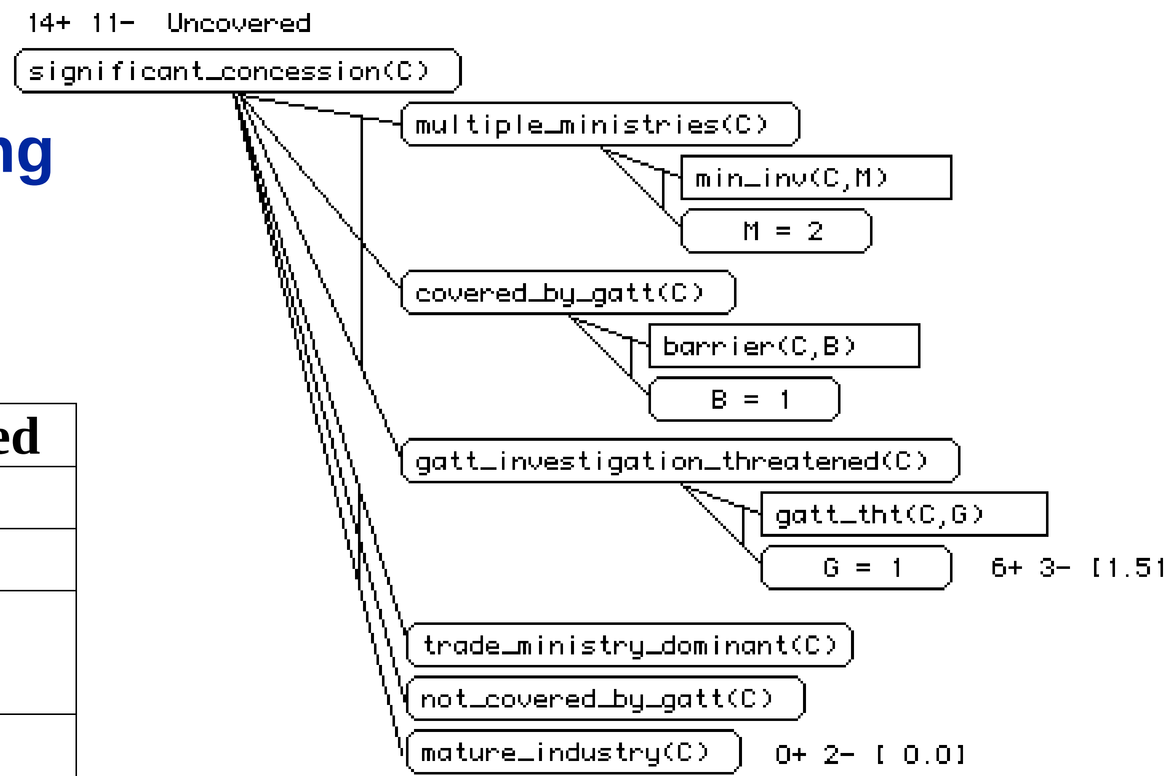
If a country has strong economic health,
Then it can tolerate a price increase.

If a country that exports a commonly available commodity tries to coerce a wealthy country, the sanction will fail because the country will buy the commodity at a higher price from another supplier

- Constrained to learning implications of existing knowledge

Theory Revision: Revising Expert Rules

- **Focus inductive learning on correcting errors in existing knowledge**
- **Search for revisions to domain theory- add or delete rules or tests from rules**
- **Experts prefer revision of expert rules to learning new rules**



Condition	Original	Revised
None	NA	68.0
Novice rules	44.0	70.0
Original expert rules	61.3	73.3
Revised expert rules	72.0	81.3

Monotonicity Constraints

- **Problem:**
  - **In some domains, experts know direction of effect of variable but not necessary and sufficient causal account.**
  - **Spurious correlations and "uninformed" selections from statistically indistinguishable tests resulted in rules that aren't understandable**
- **Monotonicity Constraints: Only use tests in intended direction**
  - **For each numeric variable: Specify if increasing values are known to increase likelihood of class membership**
  - **For each nominal variable: Specify which values are known to increase likelihood of class membership**
- **No effect on accuracy (90.7 vs. 90.6) or length (4.3 vs. 4.6) in dementia screening**

Learning a Clause with Monotonicity Constraints

Impaired 600 normal 400

Age < 68 125 150	Age < 68 100 30	Gender = F 250 5
Age < 72 170 250	Age < 72 170 40	Gender =M 200 15
Age >= 68 475 250	Age >= 68 450 20	Recall >= 2 125 18
Recall < 2 425 350	Recall < 2 375 300	Recall < 2 325 2
Months >= 2 500 50	Gender = F 275 20	Count >= 1 400 10
Months < 2 100 350	Gender = M 225 30	Count < 1 50 10

$$p_1\left(\log_2\left(\frac{p_1}{p_1+n_1}\right)-\log_2\left(\frac{p_0}{p_0+n_0}\right)\right)$$

Learning Understandable Rules for Dementia Screening

IF the years of education of the patient is > 5
AND the mistakes recalling the address is < 2
THEN The patient is NORMAL

OTHERWISE IF the years of education of the patient is > 11
AND the errors made saying the months backward is < 2
THEN The patient is NORMAL

OTHERWISE IF the years of education of the patient is > 17
THEN The patient is NORMAL

OTHERWISE The patient is IMPAIRED

Do experts prefer rules without constraint violations?

- **Procedure: generated 8 decision lists with and without monotonicity constraints (on different subsets of the CERAD)**
- **Asked 2 neurologists to rate each rule on 1-10 scale:** "How willing would you be to follow the decision rule in screening for cognitively impaired patients"
  - N1: with 5.56 without 3.25 $t(15) = 6.60$, $p < .001$.
  - N2: with 2.38 without 0.25 $t(15) = 5.09$, $p < .001$.

Correlation	**Neurologist 1**	**Neurologist 2**
Violations	.433	.623
Number of tests	.208	.020
Number of clauses	.278	.011

Learning Monotonicity Constraints

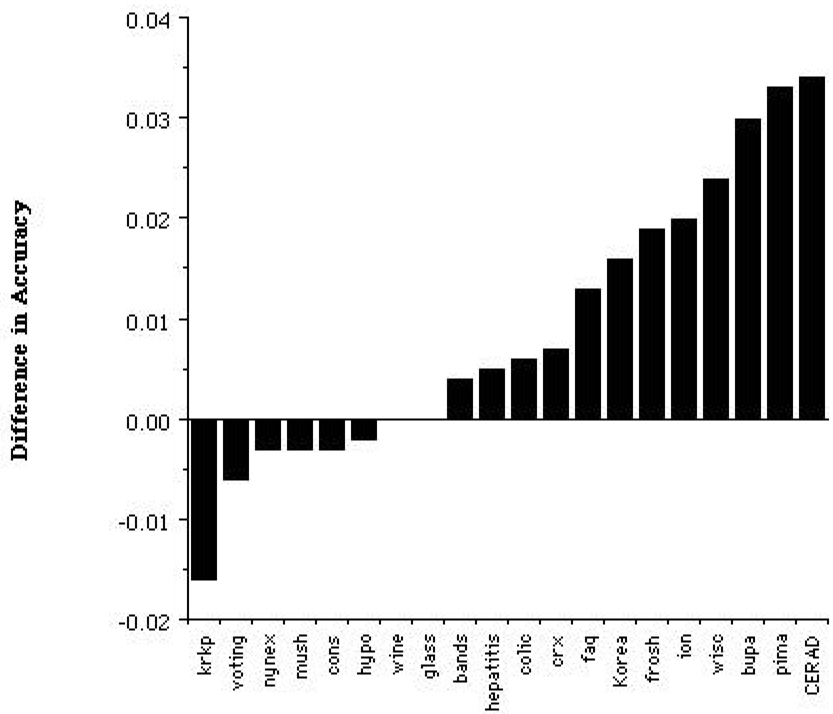
Q: Where do monotonicity constraints come from?

A: Learn them from the entire training set

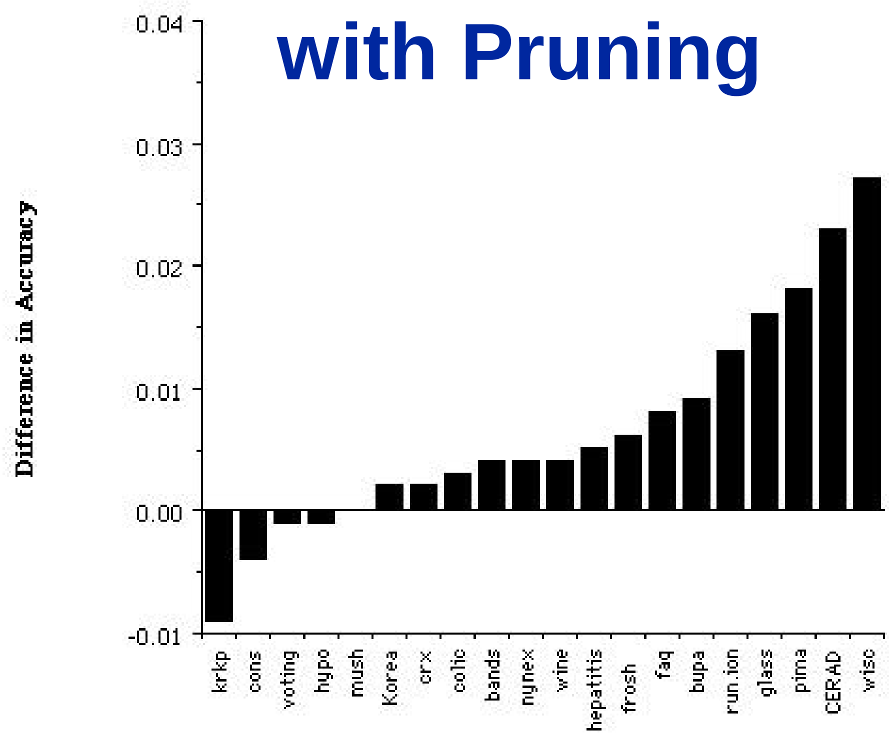
- When considering a test (selection bias)
  1. Most informative on partition of data set under consideration
  2. Informative on the entire training set
- Rationale: A variable that has the opposite effect under special circumstances is exceptional
- Disadvantage: Cannot detect negative interactions among variables.
- Preference Bias rather Selection Bias:
  - Negative interaction must be significantly superior (using chi square at 0.95 level) when used

Accuracy Results

Selection Bias

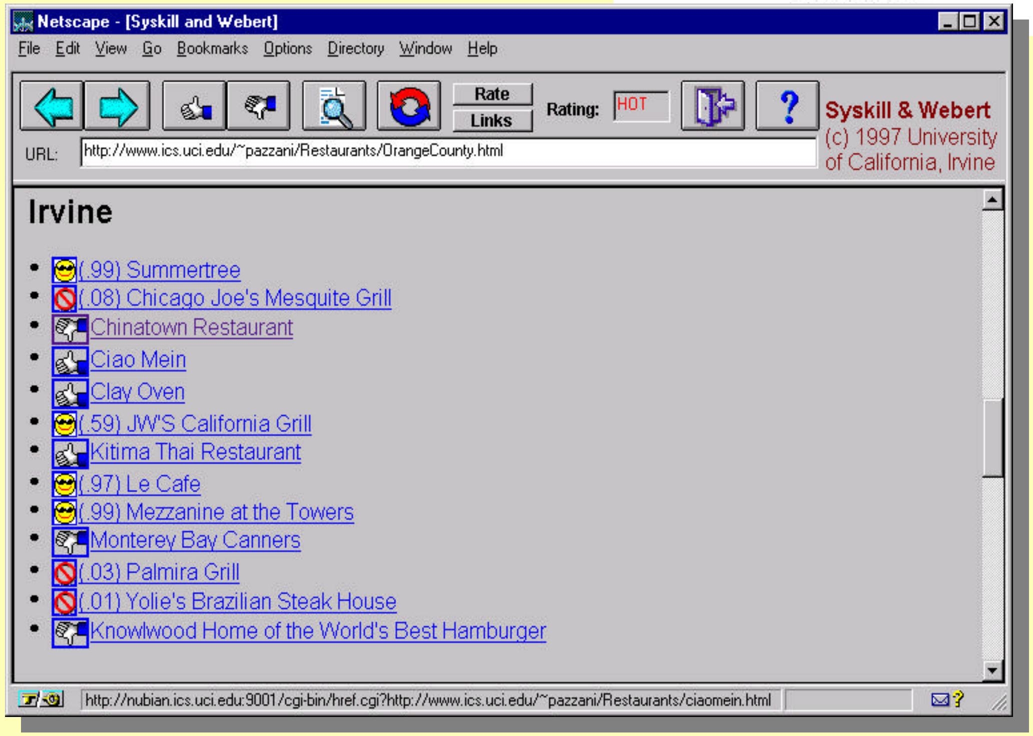


Selection Bias with Pruning



Current Research Directions

- Learning user profiles from feedback and demographics
- Explaining difference between models
  - Understand algorithms
  - Spot changes in trends
  - Identify discrepancy between specification and implementation
- Classification of time series data for intruder detection



Conclusion:
Adding knowledge to data mining gives more control over output

- To be understandable, learned concepts should conform to the cognitive biases of human experts.
- Experts prefer rules learned with monotonicity constraints.
- Current work: Explore other constraints
  - Expert judgement on learned monotonicity constraints.
  - Consistent contrast
  - Use of abstraction in concept definitions
- UCI wants your data (particularly unstructured)
  - Publicly available archive
  - Work with us under nondisclosure agreements