

Comprehensible Knowledge
Discovery: Gaining Insight from
Data by Michael J. Pazzani

Chad Roesle
EE380L - Data Mining
September 8, 1998

Purpose

- Suggest further refinements for data mining techniques
- Focus on generating comprehensible learned models, rather than just predictive models
 - Understood by experts more easily
 - Provide insight to experts
 - Support new decision-making tools

Non-comprehensible Example

- **EXAMPLE:** 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. A total of 1500 tons were delivered.

- **RULE:** 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 the product.

Comprehensible Example

- **EXAMPLE:** 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.

- **RULE:** If a country that exports a commonly available commodity tries to coerce a country with strong economic health that imports the commodity by refusing to sell them the commodity, the sanction will fail because the country will buy the commodity at a higher price from another supplier.

Rule Learning Systems

- FOIL: Inductive learning via addition of predicates to a clause body such that it encapsulates all positive examples and no negative examples
- FOCL: Extension of FOIL that utilizes a pre-defined rule base to add to or modify existing rules to be more informative

Benefits of FOCL

- Increased accuracy of the results of the learning system over FOIL
- Learned rules were only slight variations of what experts already believed
 - Easier to understand
 - Results were more readily accepted by experts
 - Experts could be retrained to use the results to obtain better predictive accuracy than before

Domain Theory

- **STRONG**
 - fairly obvious classification guidelines
 - uses a rule-base
- **WEAK**
 - rough estimations of classification guidelines
 - uses monotonicity constraints

Conclusions

- Comprehensible learned models can provide much more insight than simple inductive predictive models
- Comprehensibility and accuracy are often incorrectly considered synonymous
- Utilizing existing knowledge in learning models yields more comprehensible results