



# Methods for Prioritizing Investigations of Financial Fraud

Dr. Jianping Zhang (PI)

703-983-2839 ● [jzhang@mitre.org](mailto:jzhang@mitre.org)

Dr. Brent Han (Co-PI)

703-983-3365 ● [bhan@mitre.org](mailto:bhan@mitre.org)

Researchers: Dr. Ali Hadjarian, Dr. Jerzy Bala

Sponsor/Funding: CEM Internal Research and Development

# Problem

- ❑ **Investigation of financial fraud is a challenging task due to limited availability of resources required for in-depth examinations.**
- ❑ **There is a significant demand for predictive models to assist investigators with the task of detection and prioritization of fraud cases. These models should facilitate:**
  - ❖ **Human comprehensibility.**
  - ❖ **Ranking of suspicious cases, focusing on ones with highest risk scores.**
  - ❖ **Handling of “needle in the haystack” distributions.**

# Background

***Despite increasingly stringent legislation aimed at combating fraud, financial fraud remains a public concern.***

- ❑ Internal financial fraud increased by 19% in 2006 (Global Security Survey, Deloitte).
- ❑ U.S. organizations lost \$638 billion due to occupational fraud (2006 Report to the Nation, ACFE).
- ❑ Even auditors themselves can be active players in fraud cases (e.g., KPMG approval of accounting errors inflated revenues by \$3.0 billion and pre-tax earnings by \$1.2 billion).



# Objective

- ❑ **Develop novel rule-learning algorithms to generate human comprehensible predictive models from data sets with unbalanced class distributions.**
- ❑ **Develop effective risk-scoring algorithms.**
- ❑ **Introduce a novel approach for the optimization of the performance of developed rule-learning and risk-scoring algorithms on cases with the highest risk scores.**
- ❑ **Conduct experiments to validate the developed methods.**

# Activities

## □ Key Tasks

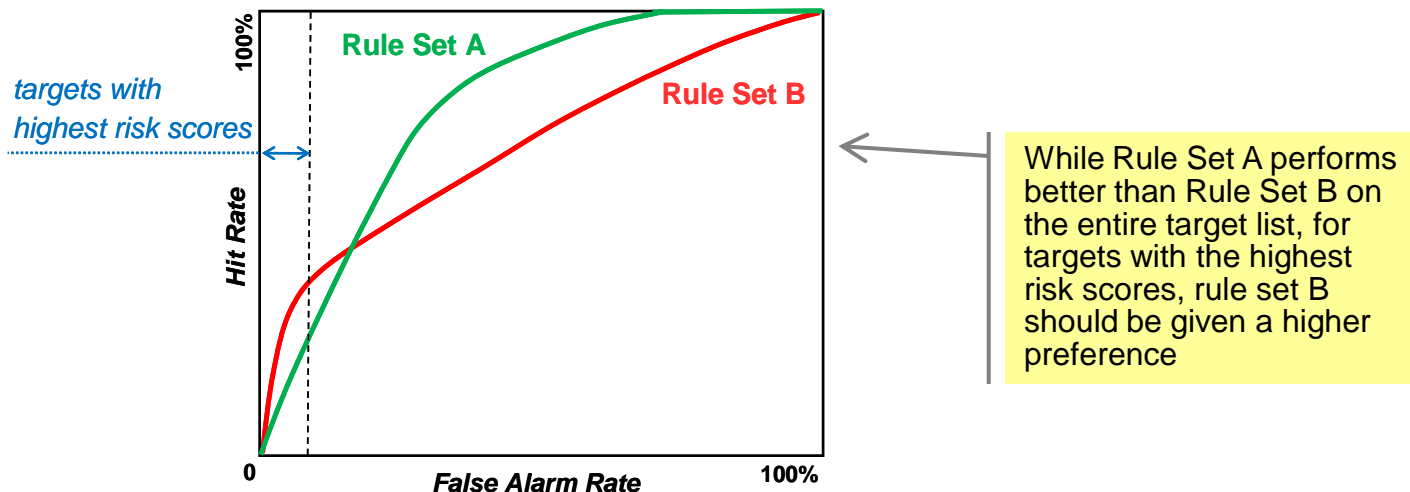
- ❖ **Survey related research in rule-learning with risk-scoring capability and learning from data with unbalanced class distributions.**
- ❖ **Design and implement novel rule-learning and risk-scoring algorithms, focusing on targets with highest risk scores.**
- ❖ **Evaluate the effectiveness of the developed algorithms on real world data.**

## □ Deliverables

- ❖ **Fully functional implementation of the developed algorithms.**
- ❖ **Scientific reports.**

# Highlights

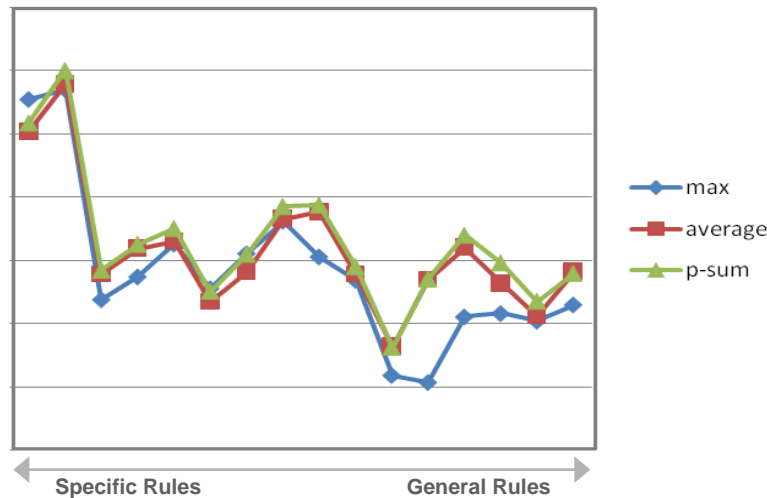
- ❑ Developed a rule-learning algorithm for data with unbalanced distributions.
- ❑ Proposed a rule-based, risk-scoring framework.
  - ❖ Methods for assigning scores using rules (e.g., probability estimation and distance calculation).
  - ❖ Method for combining scores generated by various rules (e.g., max, average, and probabilistic sum).
- ❑ Developed a novel approach aimed at optimizing a rule set's performance on targets with highest risk scores, as opposed to the entire ranked list of suspected entities.



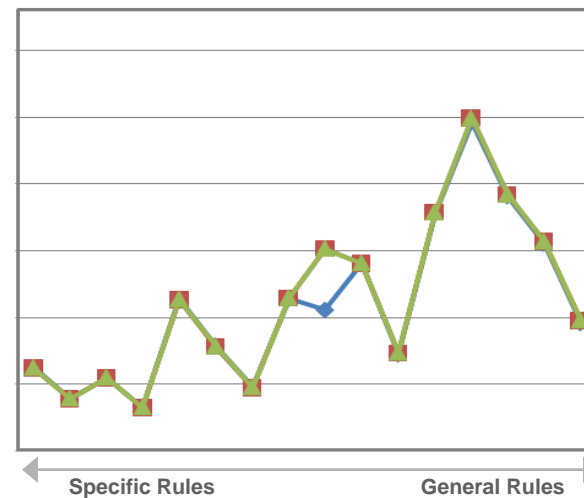
# Highlights (cont.)

- Results revealed that while certain rule-learning parameter settings can result in an increase (or decrease) in performance on the entire target list, the effect can be the opposite on the list of targets with highest risk scores.
  - ❖ Specific and precise rules work better for the targets with highest risk scores, while more general rules work better for the entire target list.
  - ❖ Three scores combining methods achieved about the same results for the entire target list, while probabilistic sum obtained better results for targets with the highest risk scores.

(a) Performance on the targets with highest risk scores



(b) Performance on the entire target list



# Impacts

## □ Customer Impact

- ❖ Provides effective methods to assist sponsors in prioritizing their investigations of financial fraud cases and optimizing the allocation of their limited resources.
- ❖ Actionable rules provide investigators with clues to look for in their in-depth examination of suspicious activities.

**Current and Potential Customers: IRS, FinCEN, SEC, PCAOB**

## □ R&D Community Impact

- ❖ Introduction of a novel paradigm for optimizing the performance of learning algorithms on top ranked cases.
- ❖ Introduction of a rule-based, risk-scoring framework.

# Future Plans

- ❑ **Continue the development of new rule-learning algorithms in line with the proposed approach for rule-based scoring, with a focus on targets with highest risk scores.**
- ❑ **Continue the development and evaluation of additional risk-scoring methods in the proposed framework.**
- ❑ **Apply the developed methods to other families of learning algorithms, including support vector machines, decision trees, and various forms of ensemble learning.**
- ❑ **Apply the developed methods to application domains other than financial intelligence.**
  - ❖ **CBP—Risk scoring for cargo screening.**
  - ❖ **DHS—Counter-terrorism.**
  - ❖ **CMS/VA—Healthcare claim fraud.**