

AI Approaches to Fraud Detection and Risk Management

*Tom Fawcett, Ira Haimowitz, Foster Provost,
and Salvatore Stolfo*

■ The 1997 AAAI Workshop on AI Approaches to Fraud Detection and Risk Management brought together over 50 researchers and practitioners to discuss problems of fraud detection, computer intrusion detection, and risk scoring. This article presents highlights, including discussions of problematic issues that are common to these application domains, and proposed solutions that apply a variety of AI techniques.

The Workshop on AI Approaches to Fraud Detection and Risk Management, held in conjunction with the Fourteenth National Conference on Artificial Intelligence (AAAI-97), was held in Providence, Rhode Island, on 27 July 1997. There were over 50 attendees, with a balanced mix of university and industry researchers. The organizing committee consisted of Tom Fawcett and Foster Provost of Bell Atlantic Science and Technology, Ira Haimowitz of General Electric Corporate Research and Development, and Salvatore Stolfo of Columbia University.

The purpose of the workshop was to gather researchers and practitioners working in the areas of risk management, fraud detection, and computer intrusion detection. We sought participants to discuss and explore common issues in the application of AI technologies to these problems, share their experiences in deploying AI approaches and techniques, and develop a deeper understanding of both the complexity of the problems and the effectiveness of various solutions. To our knowledge, this workshop was the first forum bringing together

researchers and practitioners doing work in these three related areas.

Risk management, fraud detection, and intrusion detection all involve monitoring the behavior of populations of users (or their accounts) to estimate, plan for, avoid, or detect risk. In his paper, Til Schuermann (Oliver, Wyman, and Company) categorizes risk into market risk, credit risk, and operating risk (or fraud). Similarly, Barry Glasgow (Metropolitan Life Insurance Co.) discusses inherent risk versus fraud. This workshop focused primarily on what might loosely be termed "improper behavior," which includes fraud, intrusion, delinquency, and account defaulting. However, Glasgow does discuss the estimation of "inherent risk," which is the bread and butter of insurance firms.

Problems of predicting, preventing, and detecting improper behavior share characteristics that complicate the application of existing AI and machine-learning technologies. In particular, these problems often have or require more than one of the following that complicate the technical problem of automatically learning predictive models: large volumes of (historical) data, highly skewed distributions ("improper behavior" occurs far less frequently than "proper behavior"), changing distributions (behaviors change over time), widely varying error costs (in certain contexts, false positive errors are far more costly than false negatives), costs that change over time, adaptation of undesirable behavior to detection techniques, changing patterns of legitimate behavior, the trad-

ing of accuracy for timely decisions, and social issues (privacy, discrimination, "redlining"). The following paragraphs amplify a few of these issues.

First, the probability of a bad event is extremely small in certain contexts, ranging from three to four percent of consumer credit delinquencies to fractions of one percent of fraudulent transactions.

Second, unequal, often business-driven weights are given to false positive and false negative predictions. A *false negative* means that fraud, bad credit, or intrusion passes unnoticed, with potential loss of revenue or security. However, a *false positive* means a false accusation of fraud or risk that might send away a valuable customer; lose money in challenging what is otherwise a legitimate transaction; or, in extreme cases, be the flash point for litigation.

Third, the patterns of the bad incidents change over time. For example, as fraud-detection systems become more accurate, the perpetrators invent new means of committing undetectable fraud.

The working notes contain 16 papers, 10 of which were selected for presentation at the workshop. These 10 papers were grouped into 3 categories. Four papers address issues in applying classification techniques to fraud and risk problems, including the use of clustering techniques to generate class labels (Haimowitz and Henry Schwarz of GE), the use of techniques from decision analysis and ROC analysis to deal with uncertain and changing costs and class distributions (Provost and Fawcett), and the use of metalearning techniques to address the necessity of information hiding among collaborating but competing institutions (Stolfo, David Fan, Wenke Lee, and Andreas Prodomidis, all of Columbia University, and Phillip Chan of Florida Institute of Technology).

Three papers presented approaches to modeling legitimate behavior for the detection of anomalous activity to detect fraud or intrusions. These papers addressed the problem that examples of improper behavior might be scarce and that any database of improper behavior is unlikely to be complete. Thus, patterns of proper

users must be learned and represented. Peter Burge and John Shawe-Taylor (both of Royal Holloway University) use adaptive prototypes, or statistical behavior profiles of nonfraudulent behaviors, combined with pattern-matching techniques. Partial pattern matching is also used by Terran Lane and Carla Brodley (both of Purdue University), who represent sequences of transactions for acceptable users. Jake Ryan, Meng-Jang Lin, and Risto Miikula (all of University of Texas) train backpropagation neural networks to recognize the typical commands of a computer user. An anomaly is signaled if the neural network does not classify the new commands as the actual log-on identification.

Finally, three papers look beyond currently implemented systems and prototypes for a view of what future systems should address, especially for the coming age of ubiquitous electronic commerce. In particular, systems should be able to deal with data at many different levels of aggregation (transactions, sequences of transactions, accounts). Suhayya Abu-Hakima, Mansour Toloo, and Tony White (all of National Research Council of Canada) note that systems should look beyond these aggregations to groups of accounts and transactions, for example, to identify collusive agents; they should transcend traditional systems-oriented boundaries, and they might be better viewed as investigative tools instead of stand-alone solutions. David Jensen (University of Massachusetts) discussed a government-driven prospective assessment of AI technologies for fraud detection. Henry Goldberg and Ted Senator (both of NASD Regulation) highlight "break detection" as a challenging and ever-increasing problem, requiring analysis of dynamic sequences of transactions. The papers contained a wide variety of application domains relevant to law enforcement, including cellular cloning, tumbling and subscription fraud, insurance fraud, credit card fraud, money laundering, securities fraud, check fraud, and computer intrusion.

Many issues surfaced from the papers and the workshop discussions regarding specific difficulties for building models and detectors. In particular,

the lack of existing knowledge and the existence of intelligent adversaries is problematic for building intelligent systems. Fortunately, several papers showed that machine-learning techniques have matured to a level where they can help to elicit knowledge from historical data sets and can allow systems to adapt to changing environments as new data become available from recent experience. The techniques reported on include adaptive user profiling, unsupervised learning (for example, clustering), various classification and regression models (for example, neural nets, rule learners, decision trees), link analysis, sequence matching, and fuzzy logic.

Another issue is that the decision-support scenarios needed for these problems are more complex than the simple classification problems treated in much of the AI literature. Several workshop papers address this issue by combining multiple methods. For example, Fawcett and Provost combine data mining, profiling, and classifier learning; Haimowitz and Schwarz combine unsupervised and supervised learning. Goldberg and Senator use link analysis to identify and define entities (for example, collaborators in crime). Several authors look at sequences of actions, rather than individual actions, to examine context and infer fraudulent intent.

There were also common technical issues that authors had to deal with, including lack of data on fraudulent behavior, lack of labeled data, skewed class distributions, large volumes of data, and the protection of sensitive information.

A panel discussion at the day's end focused on consumer privacy. In this era of ever-abundant online information about individuals, corporations and governments can use private information to help predict a person's credit worthiness or one's fraudulent behavior. Generally speaking, there are more restrictions on the use of demographic and personal attributes for risk management than there are for targeted marketing. David Janzen (Sprint), John Gooday (Equifax), and Barry Glasgow (Met Life) each elaborated on what information is necessary or superfluous for catching fraud and delinquency

and what information is illegal to use for risk-based discrimination. Gooday, in particular, said that sufficient attributes are available for credit scoring, but new innovations must be applied to improve performance.

Our hope was for the workshop to facilitate interaction between researchers and practitioners, focus on the commonalities among fields previously treated in isolation, and provide cross-fertilization among the fields. We thank the authors and attendees for their efforts and enthusiasm in making this possible. We are indebted to the American Association for Artificial Intelligence for organizational and funding assistance; Ray Mooney, chair of the AAAI-97 Workshop Committee; and our anonymous workshop-proposal reviewers for their suggestions and encouragement.

Tom Fawcett is a researcher in machine learning and data mining at Bell Atlantic Science and Technology in White Plains, New York. He received a Ph.D. from the University of Massachusetts at Amherst. His research interests include the effect of representation on induction, time-series problems, and the application of machine learning to real-world problems. His e-mail address is fawcett@basit.com.

Ira Haimowitz is employed at General Electric Corporate Research and Development. His research and application interests include data mining, database marketing, risk management, and automated trend detection. He received his Ph.D. in computer science from the Massachusetts Institute of Technology in 1994.

Foster Provost's research concentrates on weakening the simplifying assumptions that prevent inductive algorithms from being applied successfully. He received his Ph.D. in computer science from the University of Pittsburgh in 1992, has worked on automated knowledge discovery in science, and is currently with Bell Atlantic Science and Technology.

Salvatore J. Stolfo is professor of computer science at Columbia University and codirector of the University of Southern California/Information Sciences Institute and Columbia Center for Applied Research in Digital Government Information Systems. His most recent research has focused on distributed data-mining systems with applications to fraud and intrusion detection in network information systems. He has been awarded 9 patents in the areas of parallel computing and database inference and has cofounded two high-tech companies.