Paper 108-2013

Next Generation Detection Engine for Fraud and Compliance

Ryan Schmiedl, SAS Institute Inc.; Michael Ames, SAS Institute Inc.

ABSTRACT

SAS' next generation approach to fraud detection represents a pivotal shift in how financial institutions assess and govern customer risk. This paper discusses how companies can aggregate, sum and understand patterns in huge volumes of data; run more proactive what-if scenarios to identify and focus efforts on the most critical investigations; and understand the impacts and opportunity costs across different detection methods.

INTRODUCTION

Fraud and financial crimes continue to be a pain point for organizations across the globe. According to a 2011 survey by Mindwave Research, e-commerce merchants reported losing an average of 1 percent of total revenue to online fraud, or roughly $3.4 billion in North America. While financial institutions are typically tight-lipped with details regarding fraud losses, Discover Financial recently confirmed the growing fraud problem in their SEC10-K filing. *"We are subject to the risk of fraudulent activity associated with merchants, customers and other third parties handling customer information. Our fraud losses have been increasing and we incurred losses of $93 million, $72 million and $44 million for the years ended November 30, 2012, 2011 and 2010, respectively."* Discover also said that fraud is a growing problem for the financial services industry as well as Discover's business as it continues to expand internationally.

Financial crimes are getting more sophisticated. Fueled by a new breed of technically savvy criminals, the nature of financial crimes continues to evolve. Criminals are no longer limiting themselves to perpetrating crimes local to the institution they are exploiting. A global economy and the Internet have opened the doors for criminals to expand their fraudulent schemes to new targets. In many cases, they target an organization's technical weaknesses or process gaps. They move from one region to another quickly, and exploit any weaknesses they discover along the way.

Fortunately, there are methods to combat the growing financial crimes problem. Just as criminals leave evidence behind in physical crime scenes, perpetrators of financial crimes leave behind evidence in the form of digital evidence. In some cases, a pattern of criminal activity is pretty straightforward – for example, the structuring of deposits to bypass cash transaction reporting (CTR) requirements. In other cases, the pattern is more sophisticated as the criminal attempts to blend the fraudulent activity into normal account behavior, thus making it difficult to detect. The bottom line is that institutions need robust, flexible, adaptive detection systems to keep up with the ever-changing financial crimes challenge.

From protecting credit card use to monitoring money-laundering activities, SAS is being used to combat fraud and financial crimes by organizations across the globe. As such, SAS continues to research ways to improve its fraud detection methods. This paper outlines recent innovations SAS has been developing to improve financial crimes detection.

THE HISTORY OF FRAUD DETECTION METHODS

You can trace the lineage of modern fraud detection methods back to the emergence of rules-based and expert systems in the 1970s and 1980s. In those early days, fraud experts wrote rules for detecting patterns of behavior. However, they learned very quickly that the limited information that one could encode in a rules-based system was not enough to combat constantly changing fraud schemes. To augment the accuracy of rules-based systems, fraud experts turned to classic statistical techniques and machine learning methods to improve the accuracy of their detection schemes. The key to the success of any fraud detection system is its ability to "generalize" – that is, the ability to detect patterns of interest – in new and unseen data.

BUSINESS RULES

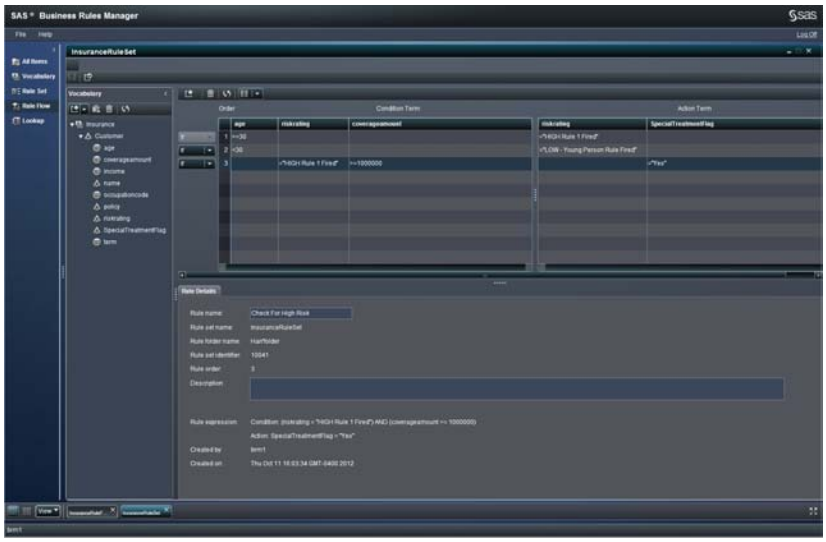
Business rules are probably the earliest and most utilized technique for detecting fraud patterns among other use cases. Rules-based approaches use structural patterns in the data to detect the desired pattern – e.g., *"IF behavior_a AND behavior_b THEN take_action."* For example, if you were writing a rule for structuring, you may write something like:

```
IF primary_medium          = "CASH"
AND transaction_type       = "DEPOSIT"
AND transaction_amount     > 10,000
```

```
THEN alert_flag     = ”TRUE”
AND alert_type      = ”CTR”;
```

Business rules have a distinct advantage in that they are both flexible and easy to understand. Most rules-based systems can be updated quickly with new rule logic without significant intervention. From a regulatory perspective, rules-based systems are typically preferred, as they can be explained easily.

To accommodate the growing need for SAS users to author and manage rules, SAS has introduced SAS® Business Rules Manager. SAS Business Rules Manager provides a complete authoring and management environment that uses the decision table metaphor as the mechanism for creating business rule logic and actions.



Display 1. SAS® Business Rule Manager

The challenge with rules-based approaches is that they encode only a small amount of information. Furthermore, rules-based approaches don't generalize well on new and unseen data because they are "information limited." The example above accounts for a situation in which a customer attempts to deposit a cash amount greater than $10,000. But consider a situation in which the customer attempts numerous smaller transactions within a single day that together total more than $10,000 or the customer purposely operates just below the $10,000 threshold? To address these patterns, the user needs to add additional rules to the detection system. This trend will continue as the organization attempts to prevent the various new flavors of fraud activity. The result or byproduct is a growth in business rules that can create management issues, overlapping business rules, and a detection engine that's always a step behind.

Despite their deficiencies, business rules have their place in fraud detection systems, as they provide a flexible, explainable method for detecting known fraud patterns. They can also be used in conjunction with other analytical detection methods. This is beneficial since business rules can be adjusted quickly, while analytical techniques typically take more time to retrain or build.

ANOYMALY/PEER GROUP ANALYSIS

Anomaly and peer grouping methods were developed to address the generalization problems of rules-based detection schemes. Peer group methods attempt to compare the behavior of individuals or entities with their peer group. A peer group is a collection of individuals or entities with shared attributes – behavioral, demographic or a combination of the two. Typically, business rules combined with clustering techniques (e.g., k-means) are used to segment populations and assign individuals and entities to specific peer groups. The behavior of an individual or entity is then compared to the peer group. Anomalies are found when an entity's or individual's behavior deviates from the "normal" behavior of the peer group. This method is popular with compliance applications, such as anti-money laundering solutions.

The challenge with peer grouping is, like rules-based approaches, its inability to generalize on new and unseen data. Another common challenge is that the behavior you are looking out for is often found to be the "normal" behavior of the peer group. And finally, there is the challenge associated with managing these types of detection systems. By gathering input from customers with these types of systems, SAS has found that difficulty managing and lack of

flexibility tend to be the primary complaints. Typical implementations use large amounts of data to create profiles within databases. This process is time-consuming, thus adaptability comes at a significant cost.

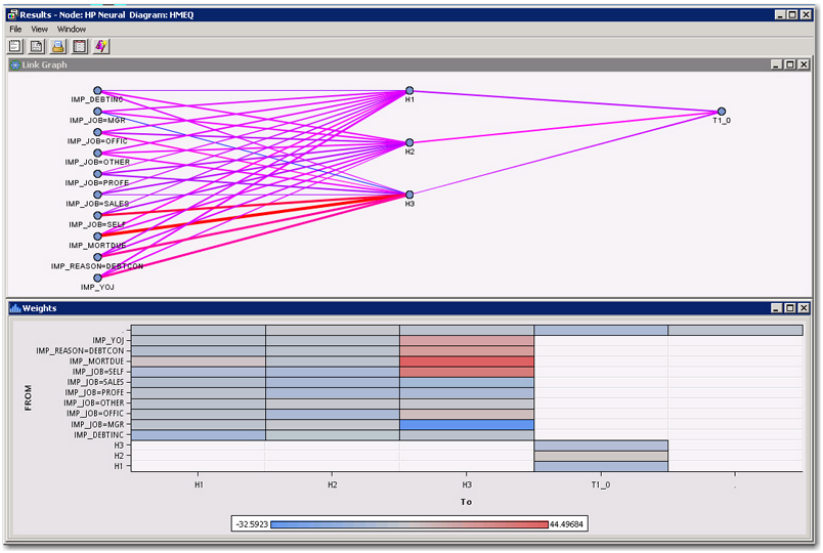
SAS is releasing a new peer group analysis method in it SAS Anti-Money Laundering solution. This new method uses the SAS® LASR™ Analytic Server to enable an interactive approach to peer group analysis. Users will be able to visualize, recalculate and – most importantly – test peer groupings on new and unseen data, thus ensuring generalizability. Using transactional base data, SAS can quickly establish new metrics and recalculate existing ones within seconds on its distributed, in-memory analytic environment.

PREDICTIVE ANALYTICS

Predictive analytics comes in a couple of different forms. Predictive analytic methods typically analyze historical data to look for patterns or relationships that provide insight into future behavior or events. A number of techniques or algorithms are utilized. A detailed explanation of each algorithm is outside the scope of this paper; however, reading recommendations are included at the end of this paper for those interested in learning more details on predictive analytical methods. Here are a few methods that are commonly used for fraud detection:

- **Decision trees** attempt to minimize classification errors by splitting the input space along decision boundaries. The decision tree partitions the data by recursively searching over candidate input variable thresholds on which to split a node and chooses the input and split point that leads to the greatest improvement in the prediction.
- **Parametric models** like logistic regression assume that the data takes on a structure that can be described by a known mathematical expression or distribution.
- **Nonparametric approaches** do not assume that the data follows any particular form. **Neural networks, K-nearest neighbors and support vector machines** are all examples of nonparametric models.
- **Ensemble methods** have been recognized as probably the most powerful methods for improving classification performance and generalizability. Ensemble methods, in their simplest form, combine the results of several models together through voting or averaging. **Gradient boosting and random forest** are two examples of decision-tree-based ensemble methods. The theory is that consensus opinions from diverse modeling techniques are more reliable than potentially biased or idiosyncratic predictions from a single source. More broadly, this principle is as basic as the saying, “two heads are better than one.”

The ability to build predictive models is at SAS’ core. SAS® Enterprise Miner™ is a keystone product that provides point-and-click predictive model-building capability.



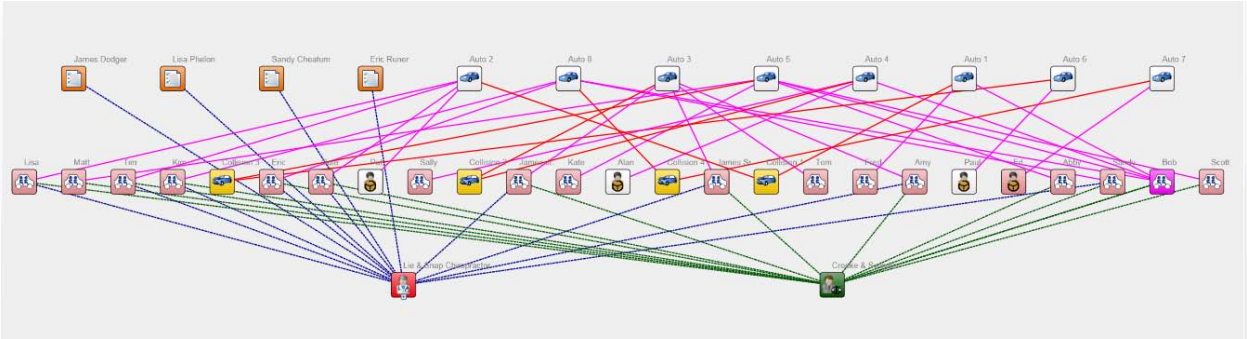
Display 2. SAS® Enterprise Miner – Neural Network with one hidden layer and 3 hidden units

The challenge with predictive analytics in fraud is threefold. One challenge is that model performance erodes over time as the events and behavior that the underlying model is attempting to predict evolve. This requires that models be retrained or new models built on more recent data. A second challenge is that the process of updating/creating new models can be time-consuming and may require resource participation from various parts of the organization.

The challenge lies in reducing or improving the processes associated with putting models into production. A third challenge involves explaining how a particular model works – in other words, a model may produce a score, but it doesn't explain how it arrived at that score. Score cards and surrogate models attempt to explain predictive models; however, they can be too complicated to understand. This is one reason that predictive models are often augmented with rules-based approaches.

LINK ANALYSIS

In its most basic form, link analysis, also known as entity resolution, establishes relationships between entities using a variety of techniques. The criteria can be as simple as exact matching (e.g., phone numbers and email addresses match) to complex fuzzy matching techniques that use complex match codes traditionally found in data quality tools for partial record matching. Statistical techniques have also been employed to identify relationships between events and entities, as well as to prune relationships to "relations of interest." Link analysis is most often used to define an entity as well as identify who may be related, how they are related and, finally, what the entity's behavior is or is likely to be. Often, link analysis is used to augment investigations by identifying nonobvious relationships and is used to extract features that can be used to enhance other detection techniques. One of the major benefits of link analysis is that it is relatively easy to combine structured content (transactions) and unstructured content (comments) into a single graph that tells a story.



Display 3. SAS® Fraud Framework – Sample Network Provider Fraud

THE NEXT GENERATION OF FRAUD DETECTION

While the previously mentioned methods have long served organizations in their efforts to prevent fraud, there are new ideas, methods and technologies that will dramatically improve fraud detection efforts. This next generation of fraud detection will use vast quantities of disparate data, distributed computing, rapid development technologies and advances in predictive modeling to produce faster, more-accurate solutions to detection problems. Here are a few guiding principles behind these next-generation methods:

- ***Use as much data as possible from a variety of sources***. Most detection systems rely on structured data in the form of transactions. It is a well-known fact that augmenting detection schemes with a wide a variety of data – including unstructured and semi-structured data – greatly improves the accuracy and generalizability of predictive models.
- ***Engineer mass quantities of diverse features from multiple sources***. Predictive performance depends on how you manipulate data and create features (variables). Next-generation systems will make it easy to assemble, manufacture and test features with more diversity and more predictive power.
- ***Explore and visualize***. Creating and selecting the features needed to detect a pattern is 80 percent of the battle. While creating simple features to perform sums, averages and counts over specific time ranges is relatively straightforward, it is much more difficult to engineer, craft and test a feature that maximizes predictive power. Success lies in the ability to create, explore and test potential features interactively on large quantities of data.
- ***Keep it simple***. If you can detect something with a simple rule, don't complicate things with models and peer groups. The minimum description length (MDL) principle says, "The solution that makes the fewest assumptions should be selected." In next-generation of detection, complex analytics will be used to engineer features and rules for detecting events of interest.
- ***Take a white-box approach to the black box***. A key limitation of current systems is the inability to provide insight into why something has been detected or how the current systems work. As the complexity of a system increases, our ability to understand the system decreases. Ultimately, what is needed from the detection engine is the story behind why something was detected.

- ***Close the loop***. Analytic systems follow a simple three-phase cycle: training, modeling and integration. As outcomes and new information are made available, they are fed back into the system to start another cycle, so that the system is refined/adapted in light of new observations and data.

CONCLUSION

Fraudsters continue to push the limits in their attempts to commit financial crimes. To prevent the next generation of even more sophisticated fraudsters, organizations must have more sophisticated detection systems. These systems must use an assortment of detection methods that take advantage of all available data. That means that new technologies will be needed to support fast, distributed processing of detection logic. These new systems will boost the productivity of users by employing various visualization methods that will help analysts gain faster insights into fraudulent schemes and more quickly identify patterns of fraud.

SAS Anti-Money Laundering uses a next-generation detection engine and new technological capabilities, including SAS LASR Analytic Server and new data visualization techniques. SAS Anti-Money Laundering can scale to previously unimaginable volumes. In addition, it can address long-standing problems with traditional AML solutions, including how to maximize investigative resources by properly setting scenario logic. Using these new capabilities, users can simulate scenarios by overlaying the outcomes of previous investigations. By visualizing the results, analysts can tune scenario parameters and understand how the inputs affect the quality of the alerts generated. The bottom line is that organizations can now focus their AML efforts in a more efficient, effective way.

Display 4. SAS® Anti-Money Laundering – Scenario Manager

Executing detection logic against large data volumes and tuning scenarios are a good start, but there is much more that can be done. If you could use all available data, alert outcomes, an assortment of analytic algorithms and a distributed, in-memory analytic environment, what else could you do? The answer is data-driven decision making. Now, how does that sound?

REFERENCES

www.cybersource.com. “2011 Online Merchants Made Most Progress Against Fraud in 13 Years.” January 14, 2012. http://www.cybersource.com/news_and_events/view.php?page_id=2173

www.discover.com. Discover Financial Services 10-K. 2012 Annual Report. January 25, 2013. http://investorrelations.discoverfinancial.com/phoenix.zhtml?c=204177&p=irol-sec&secCat01Enhanced.1_rs=11&secCat01Enhanced.1_rc=10&control_selectgroup=0

RECOMMENDED READING

- *Base SAS® Procedures Guide*
- Bolton, Richard J., and David J. Hand. 2002. “Statistical Fraud Detection: A Review.” Statistical Science. Vol. 17, No. 3: 235-255.

- Coderre, David. 2009. Computer Aided Fraud Prevention and Detection: A Step by Step Guide. Hoboken, New Jersey: John Wiley & Sons.
- Phua, C., Lee, V., Smith-Miles, K. and Gayler, R. (2005). A Comprehensive Survey of Data Mining-based Fraud Detection Research. Clayton School of Information Technology, Monash University.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Name: Michael Ames
Enterprise: SAS Institute Inc.
Address: SAS Campus Drive
City, State ZIP: Cary, NC 27513
Work Phone: (919) 531-0866
E-mail: michael.ames @sas.com

Name: Ryan Schmiedl
Enterprise: SAS Institute Inc.
Address: SAS Campus Drive
City, State ZIP: Cary, NC 27513
Work Phone: (919) 531-5421
E-mail: ryan.schmiedl@sas.com