INTERNATIONAL INSTITUTE FOR ANALYTICS™

DISCUSSION SUMMARY
RESEARCH & ADVISORY NETWORK

Analytics to Fight Tax Fraud

DEBORAH PIANKO

Government Fraud Solutions Architect, SAS Security Intelligence Practice

JUNE 2020

Interviewed by Robert Morison, IIA Lead Faculty

Discussion Overview

Tax fraud is already prevalent, and fraudsters are more sophisticated and automated than ever. To get ahead of the game in detecting fraud and protecting revenue, tax agencies need to leverage more advanced analytics including machine learning and artificial intelligence (AI). Legacy processes, systems, and attitudes need not stand in the way. To explore the challenges, opportunities, and value of tax fraud analytics, IIA spoke with Deborah Pianko, a Government Fraud Solutions Architect within the SAS Security Intelligence practice.

What are the biggest challenges facing tax agencies in the United States these days?

The most pervasive and publicized problem is identity theft. It's getting pandemic – and while the IRS says the 2016 crackdown helped slow identity theft and tax refund fraud, the IRS Commissioner said there is much more work to be done. Criminals both large and small scale, steal personal demographic and financial information and use it to file fraudulent tax returns and claim refunds. Personal information is stolen by hacking databases of consumer information, as happened at Target, or employee information, as at the Office of Personnel Management. It can also be purchased on the black market from people who have already done the stealing. And it can be stolen directly by employees at hospitals, banks, other businesses, and government agencies, including the tax agencies themselves.

Unfortunately, the IRS itself is not immune to electronic theft. Only a few weeks after the start of the 2015 income tax season, the IRS announced that their system for generating Personal Identification Numbers for online filing of tax returns had been compromised, and criminals were able to generate over 100,000 PINS. During the prior filing season, criminals stole taxpayer identity information directly from the IRS's "Get Transcript" system that provides taxpayers with copies of their tax returns and other transactions. Over a million attempts were made to access transcripts, and over half of them succeeded.

So identity theft and completely fraudulent returns are the number one challenge right now, driven by fraudsters' ability to get around the front-end detection controls in place. But there are other very common types of tax fraud and evasion. Taxpayer non-compliance includes failure to file, failure to pay, under reporting of income, and over reporting of deductions. Tax agencies want to detect all of these, but the under and over reporting cases are very difficult to notice because so much information has to be cross-referenced to verify, for example, the legitimacy of charitable deductions or the amount of sales tax charged. Because it's so hard to spot such tax evasion at filing time, agencies wind up conducting audits as much as three years later.

A major underlying challenge has to do with information systems. The internal systems that many tax agencies still run on are at least 20 years old, they were custom built, and they can't keep up with basic reporting needs, let alone detect new forms of tax fraud. But replacing an integrated tax system is a very expensive, multi-year effort ($30-50M for a typical state tax agency), and budgets are tight. Some agencies out of necessity are making the investment, and that consumes their attention at the same time that they're trying to deal with widespread identity theft and its public relations consequences.

Finally, tax agencies have a two-sided performance problem. On one side, they need to detect as much of the fraud as possible in order to minimize revenue loss. On the other side, they need to minimize the "false positives." Each time a legitimate transaction is flagged as possibly fraudulent, the subsequent investigation wastes time and effort, and it usually inconveniences the law-abiding taxpayer. There's always a tradeoff between how much fraud you want to catch and how many cases you're willing or able to work. No fraud detection system can catch all the fraud because it would generate far too many false positives. Fortunately, advanced analytics including more modern machine learning and AI techniques can help on three fronts – find fraud more accurately, flag fewer false positives, and even tune the trade-off. How many more alerts are you willing to handle to generate additional solid leads?

What's new in fraud prevention, and what does a complete capability look like?

A big trend is the adoption of identity verification quizzes. These are not the usual consumer-chosen challenge questions like "What's the first name of your maternal grandfather?" They are automatically generated "shared secrets," such as "At which of these addresses have you lived?" By the way, this is exactly the kind of protection that the IRS had in their "Get Transcript" function. It's not foolproof because the information is gathered from outside sources, including credit monitoring agencies, where fraudsters may also be able to access it. Taxpayers give these quizzes mixed reviews. They seem novel at first, a bit of a nuisance after that, and a real pain if you've forgotten some answers or the information generated is incorrect. These quizzes make tax transactions more secure, but agencies still need a second safety-net layer to detect frauds that have gotten past them.

That's just the first piece of a complete fraud prevention capability. My whole list would be:

- Identity verification system, either as a quiz or an analytically generated risk score
- Pre-payment fraud detection engine that uses both business rules and various types of analytics such as peer-group anomaly detection and predictive modeling
- Post-payment audit selection system that uses both business rules and analytics, and that predicts the likelihood of a fruitful audit
- Employee or insider threat detection capability
- Robust reporting capabilities, both regular and ad hoc, for every level of the organization
- It may go without saying, but full-force systems security and cyber security have to be in place

Those are the technological components. I'd also be sure to have a unit in the agency whose sole responsibility is revenue protection through fraud prevention. If this is divided across departments or less than a full-time focus, then accountability and controls are likely to be lax.

Tax agencies have been working on fraud prevention for years. What can they do differently and better today than they could a few years ago?

The technology has changed dramatically and with it the types of analytics we can do. Conventional tax fraud detection applications look at one return at a time and apply business rules, the equivalent of "Thou shalt not claim total itemized deductions greater than x percent of your adjusted gross income." When a rule

is broken, the return is flagged. Business rules are an effective tool to catch unsophisticated fraud, and SAS maintains a library of these types of rules for our tax fraud detection customers. But fraudsters can reverse-engineer these rules and then avoid breaking them. To keep up, the rules have to multiply and get more complex. And each return is still being evaluated in a vacuum.

The foundation of today's fraud detection is the ability to look across large numbers of returns at once and see the patterns. A single tax return claiming the Earned Income Tax Credit might not look suspicious. However, if 1,500 returns claiming the credit are filed by the same tax preparer, and all of the line items look almost exactly the same, then we're onto something. By looking across large numbers of transactions or tax returns, we can employ several forms of advanced analytics:

- With **Predictive Modeling**, we can identify suspicious returns based on the known filing patterns of proven fraudsters. For example, a fraudster might file a refund return for $1000 the first year and it's approved because it is a relatively low dollar amount and seemingly innocuous. The next year, he bumps it up and goes for $3000. The next time it's $7000. His refunds only get rejected when the amount tops $10,000. Now the fraudster has his own rule about threshold amount and he continues the scheme. Predictive analytics can spot this stair-stepping pattern being exhibited by a filer or a preparer, and score the likelihood that a given return is fraudulent or that an investigation will yield results.

- With **Anomaly Detection**, we can answer questions like, "Are this person's itemized deductions out of line with others in his income peer group or his geographic neighborhood?" or "Does the monthly variation in gross sales make sense for this type of business?" or "Does this filer's tax return seem unusual relative to prior years?" Anomaly detection uses peer grouping and builds profiles of normal and abnormal behavior. Unlike with business rules where you already have to know that you're looking for, techniques like anomaly detection automatically establish what "normal" looks like – and then automatically flag anything that looks abnormal.

- With **Link Analysis**, sometimes called "Social Network Analysis," we can see the relationships among taxpayers, tax returns, tax preparers, and others like corporate officers. People may be linked by a common address, employer, family member, bank account, and so on. As a simple example, if we have one tax return flagged for fraud, then look to see who the preparer was, we may find 100 other suspect taxpayers associated with that preparer.

Using these novel machine learning techniques allows a tax agency to let a computer do the heavy lifting instead of relying on only the gut instincts and past experiences to identify the bad guys. It can find those fishing holes that you never knew existed, instead of merely going back time and time again to the places you know the fish have historically hung out. Machine learning is the automated extraction of patterns from data. For some of these machine learning techniques, the results are captured in formulas or instruction sets so the patterns can be detected in new data, and the outcomes of those cases are fed back into the system so that the detection model "learns" and adapts over time. As fraud changes, the model changes automatically. The machines are also set up so they can alert you when your detection models are beginning to underperform. Analytical software can also create multiple detection models for the same

problem and use a "champion/challenger" process to iteratively figure out which is the best performing one. These advanced analytics are really becoming the bread and butter of tax fraud forensics.

Powerful though these analytics are, tax agencies can be very loyal to the fraud detection and audit selection business rules that have been in place for years. People are comfortable with the tools they use, and there can be a "don't worry, we've got things covered" mentality. I encourage people to spend some time with companies that build and use the new fraud detection technologies just to get educated about what's possible. Then they find it hard to imagine how they've gotten by without these capabilities for so long.

Please give some examples of how tax agencies are getting ahead of the curve. What are they accomplishing?

At one agency, analysis found a tax preparer who had submitted more than 1,900 returns in three years, more than 95 percent requesting a refund, and none of them listing a preparer. Link analysis showed that all of the seemingly unrelated returns were electronically submitted from a single IP address in increments of 15 to 30 minutes. Then, anomaly detection techniques found very large charitable contributions and job-related expenses on these returns relative to the taxpayers' peers. Without analytics, the agency's audit selection processes might have found a handful of these returns 1-3 years after the refunds had already been paid. Instead, the agency was able to pursue the preparer herself as a single criminal matter and end the scheme. When stories like this hit the press, it sends a message to the public that the Department of Revenue is on top of things, and that fraudsters aren't going to be able to fly under the radar as easily anymore.

Another state agency employed a technique called "geo-boxing." It's used by marketers in the retail industry to create microsegments or targeted neighborhoods of people who exhibit similar shopping behaviors. We helped them divide the entire state into neighborhoods of no more than 250 taxpayers whom we expected to behave similarly, and then began looking for deviations from the norm. We found a fascinating result: there were pockets where a dozen or so taxpayer returns showed exactly the same patterns, almost to the exact same adjusted gross income. Neighbors were sharing tips and tricks about how to fudge the numbers to get high refunds. I also call this "water cooler fraud," since it often happens among people who work for the same employer.

To demonstrate the value of more advanced analytics, we took a year's worth of tax returns from another state into the SAS lab. The tax refunds had long since gone out the door, but we wanted to see how much additional fraud would have been caught using analytics beyond business rules, plus more comprehensive reporting. It took only an hour for the lab to see that over $3M in tax refunds had been paid to suspect taxpayers who had filed electronically from various countries outside of the United States, including South Korea, Zimbabwe, Nigeria, and India. Now that was just for one tax year in one state.

Does adding analytics into the process slow things down, including taxpayer refunds, or speed things up?

Analytics for fraud detection is ubiquitous in the banking industry. Every time you use your credit card, rest assured that your bank is running anomaly detection models in real time to see if the transaction is consistent with your purchasing habits. With one bank, the service level agreement calls for doing that in less than 4/10th of a second for millions of transactions a day. That gives you a sense of the

amount of data that a fraud detection solution can handle and just how quickly it can provide an answer.

So detection itself happens much faster, especially compared to the overnight batch processing runs that some tax agencies still do. However, the refund cycle time can lengthen because better detection means more alerts to investigate. Agencies have used cycle time as a key performance indicator – often trying to get refunds out in seven days or less. But that happens by sacrificing fraud detection performance. Now things are shifting, and some agencies are alerting the public that they're holding more returns for longer review. The general consensus in the tax industry is that this is a change for the better. But again, to be efficient you want to control the number of false positives.

What are some of the challenges when a tax agency starts using a more analytical fraud detection system?

One of the challenges we see is that investigators need to learn how to think about alerts that were not generated by familiar business rules. If a return is flagged because the itemized deductions seem too high, that's going to be fairly straightforward for the investigator to work: send the taxpayer a letter requesting further substantiation of these deductions and wait for a response. However, if a return is flagged for a reason that sounds something like, "Taxpayer did not exhibit any seasonality in sales, though more than 90 percent of its peer group did," the investigator needs to look at new data in new ways. So there's a small learning curve in that first filing season regarding how to work novel types of fraud alerts that deal with peer grouping, behavioral grouping, and predictive modeling. But once the investigators get up to speed, they can handle more interesting cases faster. The analytics can transform their jobs.

At the organizational level, the challenge is making sure that the agency has enough people to work the alerts that are generated. The goal is to get the agency to a place where they're more efficient – reviewing fewer but better-screened alerts, or reviewing more alerts to find a lot more fraud. We're back to that tradeoff – and finding an optimal sweet spot. One agency recognized the value of the additional alerts being generated, including for later aggregate analysis, but didn't have the staff to work them. So they found a services firm to triage alerts for them, with the help of some additional analytics, so their staff could focus exclusively on alerts most likely to lead to audits.

For a tax agency that is already doing analytics as part of their revenue protection effort, can they incorporate the latest techniques without having to revamp their existing systems?

They definitely don't have to give up the legacy fraud detection models that took many years to develop and have proven fruitful up to a point. Agencies talk about tangible results, telling us things like, "We're already detecting $20M in refund fraud each year." What they don't appreciate is how much better they can be with more advanced analytics. So we recommend letting us merge our newer techniques with their set of legacy rules to see what kind of lift we can get. In one case, the approach was simply to base the decision to flag a return as fraudulent 50 percent on legacy rules and 50 percent on more advanced analytics. The hybrid and collaborative approach not only provided significant lift in generating fraud leads, but it kept the employees responsible for the old system engaged in the process as they learned new technologies and ultimately took ownership of the new system. Another agency conducted a "bake off" by running their legacy fraud detection system in parallel with a more analytically advanced one for an entire tax filing season to prove

the value of the additional lift, which was quickly in the millions of dollars.

Like fraud detection, audit selection has typically been done by business rules, and most agencies can benefit from additional analytics here as well. Which cases should be audited? How likely is an audit to be successful? And how much incremental revenue is likely to be generated? Predictive models can answer those questions, but again the approach is novel. Most agencies have a backlog of audit cases, all of which have passed the business rules screen, but no objective way to prioritize them. They also don't have good metrics and statistics on the audits they've done in the past, such as by geography, industry, and auditor. So the analytics solution builds a baseline of history to support the predictive capability of audit selection models.

How do the fraud prevention challenges and approaches differ in other countries?

Tax structures may differ widely, but for the most part, governments tax the same things – income (personal and business), consumption (VAT and sales), goods (alcohol and tobacco), and property. Fraud detection techniques and solutions are largely the same and definitely transferrable overseas.

A difficult thing overseas is the Value Added Tax (VAT). Just due to the complex way it is calculated, there are more opportunities to cheat. However, we've done VAT fraud detection projects overseas and they've been hugely successful – probably because with more opportunities to cheat, there's more fraud to find. Another challenge is that the ability of tax agencies to obtain and use data varies. In Malaysia, for example, Inland Revenue can get access to almost everything a taxpayer does at a bank. In other countries, that's against the law.

The "culture of compliance" is another big difference. In developed countries, tax obligations are generally well-understood. In developing countries, not as much, and education is just as much a lever as enforcement. But even in developed countries, there are differences. Greece and to a lesser extent Italy have a culture of non-compliance – pay as little as possible and transact in cash. That's a major factor in the financial crises faced by those countries. In contrast, Scandinavian countries have a culture of compliance, based on a strong sense of shared responsibility. The approaches we take for compliance and fraud detection should vary by local culture.

Where else can a tax agency deploy analytical solutions beyond fraud and non-compliance?

I mentioned audit selection earlier. A tax agency can use analytics to predict the likelihood of success of working any kind of inventory, whether it's fraud detection leads, suspended returns, audit cases, or collections cases. Analytics tools are so versatile. You can look for patterns and anomalies in the data to generate leads, and simultaneously to tell you which of those leads will be most fruitful to pursue — and even how to pursue them. If a tax agency is buried under unworkable inventories, especially during tax season, analytics can be used to focus and optimize their efforts.

Take collections optimization as an example. When agencies have large numbers of collections cases to pursue, but a limited number of revenue agents to work them, how do they prioritize the cases? The typical approach is to sort them by dollar amount or tax type or age. But there are variables specific to the taxpayer that influence the likelihood of successful collection, starting with location and assets. Queuing by dollar amount will seldom yield the optimal result. With analytics, an agency can prioritize by likelihood the overall success and predict expected revenue.

Analytics can also recommend customized collections approaches for each individual case based on the characteristics of taxpayer and debt. What set of treatments will build up just the right amount of pressure to get the taxpayer to pay? This is what private debt collectors do, so why shouldn't tax agencies be using the same techniques? Optimization is about working smarter – using the resources you have to generate more revenue than you can today.

What are agencies' blind spots when it comes to reducing tax fraud?

I'll give you two. Tax agencies may assume that their integrated tax systems are already doing analytics, especially if the system is relatively new and very expensive. But integrated tax systems tend to have limited analytical capability. They excel at what they're designed to do — automate returns and payment processing, issue bills and refunds, and manage audit and collections cases. I encourage tax administrators to learn about what analytics can do, assess how much the existing system is really capable of, and supplement with additional analytics as needed. In all of the states where we support tax fraud detection, the advanced analytics coexist or integrate with the integrated tax systems. Don't let your system or its vendor be the gatekeeper between you and more modern technologies and techniques.

The second common blind spot is ignoring inside fraud. EMTs are taught never to get the "big eye," which is the fire department's way of saying, "Don't get so obsessed with an injury bleeding a lot that you miss the fact the person is also having a heart attack." The fraud detection version is, "Don't be so preoccupied with identity theft and taxpayer fraud that you fail to notice what employees are doing." Are any of them committing tax refund theft or helping their friends and family get their tax liabilities adjusted downwards?

In one extreme case, a manager in a state agency's property tax division stole $50M in tax refunds over 15 years. She started small and, once successful, took larger and larger property tax refunds in the names of fake businesses "owned" by friends and relatives. Basic reporting was so poor that the external auditors couldn't see the pattern. The fraud was finally exposed thanks to a bank employee who sensed something fishy about a refund check presented for deposit.

Most agencies would have to sift through all the financial adjustments made by employees to notice which ones don't look right. That's impossible in a large office that processes hundreds, if not thousands, of adjustments and refund approvals per day. Even then, you'd notice isolated cases, not patterns of fraud. However, the same analytics tools and methods used for taxpayer fraud can be applied to inside workflows. Given the financial and reputational risk of having just a single employee or group perpetrating fraud, tax agencies can't afford *not* to use analytics as a monitoring tool.

To summarize, what are the top three things that tax agencies should know and do about analytics for tax fraud prevention?

Here are a few quick things I always tell people, and I think we've touched on all three:

- First, don't get that "big eye" around identity theft on income taxes, prevalent and problematic though it may be. Don't let fraud detection in other taxes, or non-compliant taxpayers, or insider threats, fall to the wayside.

- Second, make sure you have great analytical and statistical reporting, and that investigators and other staff can generate their reports directly. Give them the information

access and tools to act on alerts quickly, to investigate deeply, and to explore the patterns they see.

- Third, and most importantly, don't assume that your integrated tax system is already doing analytics because it has a "fraud module." Find out what's really under the hood. Advanced analytics solutions — and those for tax fraud detection in particular — are inexpensive relative to your core tax system, they can be implemented quickly, and the two systems can work together to reduce fraud and protect revenue.

Additional Information

To learn more about this topic, please visit www.sas.com/taxfraud.

About the Interviewee

DEBORAH PIANKO

Deborah Pianko is a Government Fraud Solutions Architect within the SAS Security Intelligence practice working with state and local governments to combat waste, fraud and abuse. Deborah has close to 20 years of experience building technology solutions for tax and revenue agencies. She is a subject matter expert in tax administration including collections, audit, return and payment processing, customer service, revenue protection and fraud detection and prevention.

Prior to joining SAS, Deborah was a consultant to the District of Columbia (DC) Office of Tax and Revenue where she served as the lead forensic data miner and analyst during the investigation of two criminal refund fraud cases involving DOR employees, and one external-filer refund fraud case involving the use of prepaid debit cards. She worked with the local government, FBI, and external auditors to help the District successfully prosecute the cases and earn a clean opinion during the annual audit. She also was the lead data miner on a project designed to expose Medicaid fraud by comparing aid recipient data to the tax rolls.

Deborah also led the introduction of a Visa Prepaid Debit Card program for tax refunds which successfully went live in January 2014. She also served as the lead project manager for DC's 2010 Tax Amnesty program. This included billing 45,000 taxpayers for $150M of accounts receivable, processing all incoming payments, and evaluation of the 45,000 accounts for amnesty benefits. DC surpassed its goal of collecting $20M through this program.

Deborah spent many years as a computer programmer and business analyst in the tax and revenue industry and has helped to build and maintain many of the integrated tax systems that exist at departments of revenue across the country today such as DC, Tennessee, Arizona, Maryland, Ohio, Puerto Rico, Detroit, and others. Her intimate knowledge of these systems allows SAS' tax and revenue clients to hit the ground running as they make their leap forward towards a new generation of fraud solutions.

Deborah holds a dual Bachelor of Science degree in Economics and Journalism from Rutgers University. She and her husband, Curt, reside in the suburbs of Washington, DC with their son and dog. In her spare time, she is an avid rower and a volunteer EMT/Firefighter.