Adding ‘augmented intelligence’ and ‘data visualization’ to the mix

Augmented intelligence and data visualization are two technologies fraud examiners can use to make the analysis phase of a fraud examination more efficient. Augmented intelligence through “predictive coding” allows fraud examiners to use their unique knowledge and intuition to identify and evaluate relevant documents and data.

And CFEs can use data visualization (graphs or other visual presentations) so that most non-financial people (including judges and juries) can understand the nature, trends and meaning of complex financial data instead of trying to decipher information in traditional tabular form.

Augmented intelligence through predictive coding

While artificial intelligence and machine learning (computer systems that can learn from data unassisted — a subset of artificial intelligence) are becoming more commonplace concepts, the term “augmented intelligence” still causes confusion. This story, told by data intelligence agent Shyam Sankar in his June 2012 TED Talk, best defines this concept:

An international chess tournament was held in 2005, in which contestants could choose any combination of man and machine desired, and any level of skill was allowed. Team combinations included grand masters with laptop computers, and supercomputers, among others. Ultimately, two amateurs using three average laptop computers won the tournament, defeating grand masters and supercomputers alike. It wasn’t their chess-playing skill that allowed them to win, nor their programming skills — it was the ability to effectively work with their computers that made the difference. (Watch the video of his talk, “The rise of human–computer cooperation,” at http://tinyurl.com/mqtno2c.)

The essence of “augmented intelligence” is the unification of the best human skills with the best advantages of machines.

Predictive coding, simplified

The legal profession uses the term, predictive coding (sometimes called “technology-assisted review”) to describe a technology that can search through vast collections of documents and retrieve matches by both keyword and concept. Outside the legal world, this might be called “concept searching” or something very similar to Google’s “find more like this.” Here’s how it works in a non-litigation investigation setting:

1. The investigation team identifies an initial set of relevant documents or emails.
2. The team passes the relevant set through artificial intelligence-based machine learning algorithms, which employ natural language processing (NLP) functions to search for conceptually similar documents.
3. The team reviews the resulting set of potentially relevant documents and again assesses their relevance.

4. The iterations continue until the team reaches a reasonable comfort level. Each pass returns more relevant material and also uses the “not relevant” material to validate the results.

The purpose of predictive coding in a fraud examination is to find relevant content faster and with greater accuracy than traditional methods such as keyword searches or manual review. Often a CFE will use it early in an examination to help “widen the net” more efficiently or near the end of an examination as a “final sweep.” The fraud
examination use of this technology differs markedly from the more restrictive, formal approach that’s necessary in a litigation environment.

The best of us, the best of “them”
During fraud examinations, CFEs determine case theories and relevancy plus incorporate those intangibles — critical experience, discernment and hunches — that “expert systems” (a computer system that, through decision rules, emulates the judgment of a human expert) can’t successfully emulate. Conversely, fraud examiners can suffer from mental fatigue during prolonged hours of reading documents. As a 2012 RAND Corporation study put it, “…human reviewers exhibit significant inconsistency when examining the same set of documents for responsiveness under conditions similar to those in large-scale reviews. …” (See the RAND Institute for Civil Justice PDF, “Where the Money Goes: Understanding Litigant Expenditures for Producing Electronic Discovery” by Nicholas M. Pace and Laura Zakaras, http://tinyurl.com/8zxxcez.)

So, machines can augment the fraud examiner’s skills. Modern NLP systems are quite adept at retrieving conceptually similar documents from a large collection in less time than their human counterparts. Published research has shown this method to be superior to both human-only review and traditional methods such as keyword search. (See the PDF, “Technology-Assisted Review in E-Discovery Can Be More Effective and More Efficient than Exhaustive Manual Review,” by Maura R. Grossman and Gordon V. Cormack, Richmond Journal of Law and Technology, Vol. XVII, Issue 3, http://tinyurl.com/3vn6h8z.)

The real benefits
While the hype around this technology focuses on finding the “smoking gun” evidence, the primary benefit is about what documents you don’t review. Conservative estimates put the reduction in amount of text items to review (and possibly even more important, cost) at 80 percent. (See “E-Discovery And The Rise of Predictive Coding” by Ben Kerschberg, Forbes, March 23, 2011, http://tinyurl.com/m3hbgg3.) However, our experience has shown this reduction to be even higher. On a recent investigation we worked on, which involved 2.5 million emails, predictive coding reduced the volume of potentially relevant materials by more than 99 percent. In the end, there were only 250 emails relevant to the investigation.

Suited for academics
Because of the complexity and enormous cost of current eDiscovery predictive coding tools, this technology hasn’t found its way into mainstream use in the fraud examination and forensic investigation worlds. To exacerbate this, the legal profession has necessarily installed a cumbersome set of processing rules that govern the process from start to finish. Attempts to create and use alternative tools (such as open source) require knowledge of NLP and linguistics, programming, investigative principles and domain expertise. This can be a daunting task for corporate entities and nearly impossible for individual practitioners. Yet, the need for predictive coding in fraud examinations and forensic investigations has never been greater. Academic institutions are uniquely suited to answer this call — nowhere else on earth is there a concentrated collection of experts in the necessary areas.

Data visualization
With the rise in popularity of the infographic (visual representations that quickly communicate often complex facts and figures) it’s easy to confuse the concept of data visualization with data art. Simply put, data visualization is the process of creating a visual representation of data. For the fraud examiner, the two main benefits are enhancing the analysis of data and clarifying the explanation of data-heavy conclusions.

Analyzing relationships, trends and patterns of transactions, sequences of events and shifts in patterns over time are all complex processes. Attempting to make sense of these items using only the data can be cumbersome — one that may not show you what others truly need to see. Data visualization helps solve this challenge.

Relationships
A one-to-one relationship between a vendor and an employee doesn’t take much to comprehend. However, when analyzing thousands of vendors and hundreds — if not thousands — of employees, it can be much more difficult to comprehend the significant relationships. Data visualization helps illustrate and indicate the key relationships, nodes of activity and unusual relationships very quickly. (We’ll discuss the process of analyzing these relationships further in our next column.)

Trends and patterns in transactions
A common red flag in a conflict of interest, related party or kickback scheme is an acceleration pattern of activity for the involved vendor. While the data will contain the dollar figures and this pattern of activity, it’s hard to see the pattern when looking solely at the data. Fraud examiners, using data visualization, can view...
the trends on a number of different time scales, including daily, weekly, monthly, quarterly or annually.

**Sequences of events**
As we discussed in a previous column, the unstructured data elements discovered using digital forensics typically contain date and time elements. [See “The leading edge of high tech: The collaboration between data mining and digital forensics,” in the July/August issue of Fraud Magazine, at http://tinyurl.com/k9ahacs.] Plotting key events and activities on a chronological scale that encompasses transactional activity can help illustrate correlation between key events and individuals. The examiner then can highlight, organize and drill down into the individual activities to determine significance.

**Application to academia**
The conveyance of critical details and facts become the key drivers of the design of data visualizations. As stated by Nathan Yau, Ph.D., in “Data Points: Visualization That Means Something” (http://tinyurl.com/p7enez5), a good visualization is “…a representation of data that helps you see what you otherwise would have been blind to if you looked only at the naked source.”

Individuals in their academic careers are in unique positions to learn the foundational knowledge necessary to excel in the area of data visualization. Unlike a traditional accounting role, an individual who excels in this area has knowledge of statistics, graphic design, communications and, many times, accounting. Those with the ability to combine these specialty areas into useful analysis tools and media will continue to be in greater demand at higher-education institutions. ■ FM

Les Heitger, Ph.D., Educator Associate, is BKD Distinguished Professor of Forensic Accounting in the School of Accountancy at Missouri State University in Springfield. He’s chair of the ACFE Higher Education Advisory Committee. His email address is: LesHeitger@MissouriState.edu.

Jeremy Clopton, CFE, CPA, ACDA, is senior managing consultant in the Forensics Practice of BKD, LLP. His email address is: jclopton@bkd.com.

Lanny Morrow, EnCE, is a managing consultant in the Forensics Practice of BKD, LLP. His email address is: lmorrow@bkd.com.