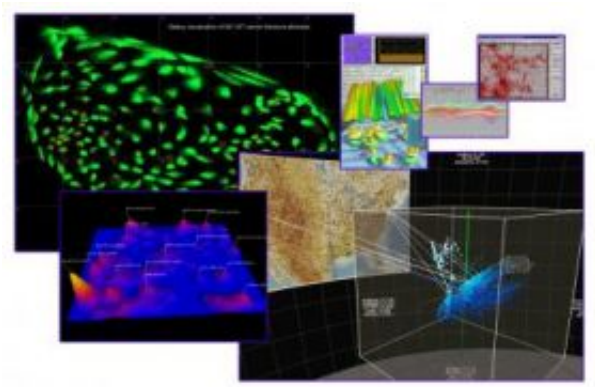
Google meets Sherlock Holmes

November 14 2007



Scatter plots, 2-D shapes, rotatable 3-D clouds, animation -- these and other techniques are being explored to help analysts see clues in mountains of "fuzzy" data. Credit: DHS S&T

Soon after 9-11, Americans wondered aloud: How did our guardians miss the clues? Suspects on watch lists had moved money in curious ways. “Chatter” had risen in recent months. A visitor to the country had offered cash to learn how to fly—but not land—a jetliner. In hindsight, these telltale nuggets provided evidence of the terror to come.

Or did they? Most such nuggets were buried in a landslide of data arriving faster than analysts could make sense of it. A day’s take would fill more than 6 million 160-gigabyte iPods. Moreover, like people, the nuggets sometimes disagreed. And like a story told and retold, their message changed, sometimes imperceptibly.

Finally, most nuggets are cast in unstructured, “fuzzy” data. The same face—or is it"—may appear in three surveillance videos. Someone in Florida is snapping up potential makeshift detonators on eBay. Such clues, like most, don’t come conveniently packaged in a tidy spreadsheet or searchable text; they must be inferred from photos, videos, voice.

To thwart the next 9-11, analysts must meld the encyclopedic eye of Google with the inductive genius of Sherlock Holmes.

Late last century, Edward Tufte catalogued ways to display data that were either structured (train schedules) or similar (death rates). Today, researchers at the DHS Science and Technology Directorate are creating ways to see fuzzy data as a 3-dimensional picture where threat clues can jump out.

The field of visual analytics “takes Tufte’s work to the next generation,” says Dr. Joseph Kielman, Basic Research Lead for the Directorate’s Command, Control and Interoperability Division. Kielman advises the National Visualization and Analytics Center, based at Pacific Northwest National Laboratory, and its university partners, called the regional centers.

The centers’ interdisciplinary researchers are automating how analysts recognize and rate potential threats. Mathematicians, logicians, and linguists make the collective universe of data assume a meaningful shape. They assign brightness, color, texture, and size to billions of known and apparent facts, and they create rules to integrate these values so threats stand out. For example, a day’s cache of video, cell phone calls, photos, bank records, chat rooms, and intercepted emails may take shape as a blue-gray cloud (picture, lower-right). If terror is afoot in L.A. and Boston, those cities are highlighted on a U.S. map (picture, center).

A month of static views might be animated as a “temporal” movie, where a swelling ridge reveals a growing threat.

“We’re not looking for ‘meaning,’ per se,” Kielman explains, “but for patterns that will let us detect the expected and discover the unexpected.” Neither the researchers nor the analysts, he says, need to understand the terrorists’ language—no small advantage, given the shortage of cleared linguists.

It will be years before visual analytics can automatically puzzle out clues from fuzzy data like video, cautions Kielman: “The pre-9/11 chatter didn’t say, ‘We’re going to plow airplanes into the Twin Towers.’ To correlate these facts, you must get relational,” connecting screen names with bank records, bank records with faces. How researchers will get there remains an unwritten story. But with each chapter, the plot thickens.

Source: US Department of Homeland Security