

Inventing the 'Google' for predictive analytics

December 20 2017, by Rob Matheson



An illustration of real-world behavioral commonalities in raw data of transactions. Credit: Endor

Companies often employ number-crunching data scientists to gather insights such as which customers want certain services or where to open

new stores and stock products. Analyzing the data to answer one or two of those queries, however, can take weeks or even months.

Now MIT spinout Endor has developed a predictive-analytics platform that lets anyone, tech-savvy or not, upload raw [data](#) and input any business question into an interface—similar to using an online search engine—and receive accurate answers in just 15 minutes.

The platform is based on the science of "social physics," co-developed at the MIT Media Lab by Endor co-founders Alex "Sandy" Pentland, the Toshiba Professor of Media Arts and Sciences, and Yaniv Altshuler, a former MIT postdoc. Social physics uses mathematic models and machine learning to understand and predict crowd behaviors.

Users of the new platform upload data about customers or other individuals, such as records of mobile phone calls, credit card purchases, or web activity. They use Endor's "query-builder" wizard to ask questions, such as "Where should we open our next store?" or "Who is likely to try product X?" Using the questions, the platform identifies patterns of previous behavior among the data and uses social physics models to predict future behavior. The platform can also analyze fully encrypted data-streams, allowing customers such as banks or credit card operators to maintain data privacy.

"It's just like Google. You don't have to spend time thinking, 'Am I going to spend time asking Google this question?' You just Google it," Altshuler says. "It's as simple as that."

Financially backed by Innovation Endeavors, the private venture capital firm of Eric Schmidt, executive chairman of Google parent company Alphabet, Inc., the startup has found big-name customers, such as Coca-Cola, Mastercard, and Walmart, among other major retail and banking firms.

Recently, Endor analyzed Twitter data for a defense agency to detect potential terrorists. Endor was given 15 million data points containing examples of 50 Twitter accounts of identified ISIS activists, based on identifiers in the metadata. From that, they asked the startup to detect 74 with identifiers extremely well hidden in the metadata. Someone at Endor completed the task on a laptop in 24 minutes, detecting 80 "lookalike" ISIS accounts, 45 of which were from the pool of 74 well-hidden accounts named by the agency. The false positive rate was also extremely low (35 accounts), meaning that human analysts could afford to have experts investigating the accounts.

Clusters of commonality

Machine learning is used for complex computational problems that are relatively static, such as image recognition and voice recognition. Written and spoken English, for instance, has been essentially unchanged for centuries.

Human behavior, on the other hand, is ever-changing. Predicting [human behavior](#) means analyzing a large number of small signals over a short period of time, perhaps days or weeks. Traditional machine-learning algorithms rely mainly on constructed models that analyze data over much longer periods.

"In general, you need a lot of data to build accurate models for human behavior, and that means you have to rely on the past. Because you rely on the past, you cannot detect things that recently happened, and you can't predict human behavior," Altshuler says.

Throughout the early- and mid-2000s, Pentland and Altshuler developed "social physics" in the Human Dynamics Lab, with aims of capturing and analyzing short-term data to understand and predict crowd dynamics. In their research, they found all big data contain certain mathematical

patterns that indicate how social interactions spread and converge, and those patterns can help predict future behaviors.

Using those mathematical patterns, they built a platform—the core technology of Endor's platform—that can extract "clusters" of behavioral commonalities from millions of [raw data](#) points, much more quickly and accurately than machine-learning algorithms. A cluster may represent families of four, people who buy similar foods, or individuals who visit the same locations. "Most of those data patterns would be indistinguishable from noise with any other technologies," Altshuler says.

It isn't immediately clear what clusters represent, just that there is a strong correlation. Querying the data, however, provides context. With customer data, for instance, someone might query which customers are most likely to buy a specific product. Using keywords, the platform matches behavioral traits—such as location and spending habits—of customers who have bought that product with those who haven't. This overlap creates a list of possible new customers that are apt to buy the product.

In short, uploading data and asking the right question presents the platform with a basic request: Here is an example X, find me more of X. "As long as you can phrase a question in that way, you'll get an accurate response," Altshuler says.

Endor and Endor-ish

To test the platform, the researchers worked early on with the U.S. Defense Advanced Research Project Agency (DARPA) to analyze mobile data in certain cities in times of civil unrest to show how emerging patterns can help predict future riots. Altshuler also spent a couple months in Singapore analyzing taxi ride data to predict traffic

jams in the city.

In 2014, Altshuler connected with Schmidt through Doron Alter, a friend and Stanford University graduate, who at that time was a partner in Innovation Endeavors. The investors asked if the technology could be wrapped "into a product that could be used by anyone," Altshuler says.

That year, with Schmidt's financial support, Altshuler and Pentland, a serial entrepreneur, co-founded Endor to transform the platform into commercial software. The team was joined by Alter and Stav Grinshpon, a tech-industry veteran and former leading technical expert at 8200, an Israeli Intelligence Corps. unit.

The company had soon earned an early partner in Mastercard through the credit card company's StartPath program. Altshuler was asked by Mastercard to answer queries reserved for data scientists, such as who is going to fly abroad soon, take out loans, or increase credit card activity.

On a single flight from Tel Aviv, Israel, to New York City, Altshuler crunched billions of data points on financial transactions of 1 million card-holders and received accurate answers to 10 questions.

Traditionally, data scientists would need to spend weeks, or months, cleaning the data and designing machine-learning models to answer each question individually. "It would have taken the company, say, two months to develop models to answer those questions. I did 10 on one transatlantic flight," Altshuler says.

Companies may employ their own analytics-savvy staff to use Endor. Others will set up brief weekly meetings with Endor representatives to determine the best phrasing for questions. "It takes about five minutes to translate their English to what we call 'Endor-ish,' meaning the way our system can understand questions," Altshuler says.

The startup's webpage offers an example of results and a comparison with traditional machine-learning engines. A marketing department for a bank asks, "Who is going to get a mortgage in the next six months?" Machine-learning engines may detect a pool of, say, 5,000 customers who have a bank [credit card](#) and a high credit score, and are married—many of which may be false positives. Endor detects more specific clusters of, say, couples about to get married or going through a divorce, founders who recently sold their startups to Facebook, or customers who recently graduated from a local real-estate course. Results from Endor offer far fewer false positives and dig up far more additional potential customers, according to the startup.

Importantly, Altshuler says, Endor isn't aimed at replacing data scientists; it's designed as a tool to empower them. Data scientists, he says, are most familiar with their organization's business semantics and can incorporate Endor into their workflow. By opening a "bottleneck"—where data input comes in faster than anyone can produce an output—Endor aims to help data scientists improve their companies. "Data scientists understand we can make them heroes," Altshuler says.

Endor recently won the "Cool Vendor" status by Gartner, reserved for industry disrupters, and was acknowledged as a "Technological Pioneer" by the World Economic Forum. As word spreads, Endor is now gaining customers across the U.S., with first customers also in Europe and Latin America. "It's exciting times," Altshuler says.

This story is republished courtesy of MIT News (web.mit.edu/newsoffice/), a popular site that covers news about MIT research, innovation and teaching.

Provided by Massachusetts Institute of Technology

Citation: Inventing the 'Google' for predictive analytics (2017, December 20) retrieved 10 May

2024 from <https://phys.org/news/2017-12-google-analytics.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.