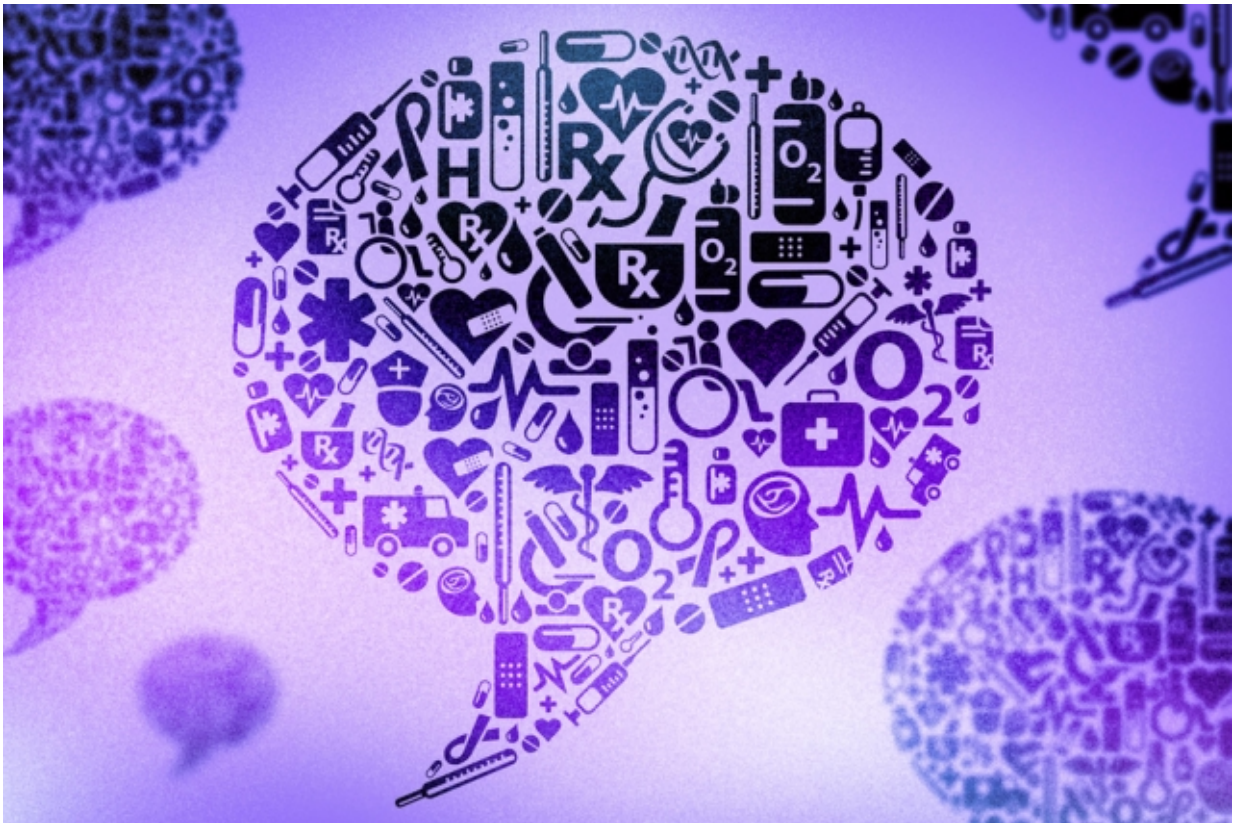


# Detecting consumer decisions within messy data

December 22 2015, by Rob Matheson

---



“In health care, there’s this gigantic world of unstructured data that needs to be translated into useable information,” says Paul Nemirovsky, who co-founded dMetrics with Ariadna Quattoni. Credit: MIT News

Millions of people each month report positive and negative health care

feedback across the Web. Some jump into forums to complain about ineffective prescriptions or to discuss which drugs are best to treat illnesses. Others take to blogs to describe symptoms and how to get relief.

MIT spinout dMetrics believes this online chatter is an information treasure-trove for the health care industry. "In health care, there's this gigantic world of unstructured data that needs to be translated into useable information," says Paul Nemirovsky PhD '06, who co-founded dMetrics with Ariadna Quattoni PhD '09.

The startup has developed a platform called DecisionEngine that uses machine learning and natural language processing—which helps computers better understand human speech—to mine billions of conversations about drugs, medical devices, and other health care products. These discussions are happening on blogs, Facebook, Twitter, forums, and even in comments accompanying news articles and videos.

From those vast stores of messy data, the software reveals insights into consumer decisions, Nemirovsky says: "What people do, don't do, consider doing, may do, did in the past, as well as what needs, fears, and hopes they have."

Today, Nemirovsky explains, dMetrics has a database that includes every public comment about patient-reported illnesses, solutions, and outcomes, pulled from more than 1 million online sources. This includes information on more than 14,000 health care products.

Clients, including Fortune 500 companies and nonprofit organizations, can use dMetrics software to answer specific questions, such as how many patients used a specific medication for a particular reason in certain time frame, or which customers are considering switching from their drug to a competitor's drug.

Although focusing on the health care industry, dMetrics, headquartered in Brooklyn, New York, is also trialing its platform with consumer finance and political organizations. Credit card companies, for instance, can analyze why consumers favor specific credit cards over others. Political scientists could use the software to determine which issues people care about and how strongly they stand behind their opinions.

"For all these types of questions, you have to understand not only the words people use but the concepts behind the words," Nemirovsky says.

## **Decoding language and expression**

Other software generally relies on ontologies—formal naming and definitions—to sense overall sentiment and popularity of brands, Nemirovsky says. The software may count, for example, the number of mentions of a word (such as the name of a specific drug) to determine if it's important, or it may detect "positive" or "negative" words.

"Language and expression doesn't work like that," Nemirovsky says. "We're a bit more complex as humans."

DecisionEngine, Nemirovsky says, better derives meaning from text because the software—which now consists of around 2 million lines of code—is consistently trained to recognize various words and synonyms, and to interpret syntax and semantics. "Online text is incredibly tough to analyze properly," he says. "There's slang, misspellings, run-on sentences, and crazy punctuation. Discussion is messy."

Visualize the software as a three-tiered funnel, Nemirovsky suggests, with more refined analysis happening as the funnel gets narrower. At the top of the funnel, the software mines all mentions of a particular word or phrase associated with a certain health care product, while filtering out "noise" such as fake websites and users, or spam. The next level down

involves separating out commenters' personal experiences over, say, marketing materials and news. The bottom level determines people's decisions and responses, such as starting to use a product—or even considering doing so, experiencing fear or confusion, or switching to a different medication.

To explain, Nemirovsky provides an example comment that could appear in an online forum: "I'm now on Drug A and took 10 mgs of Drug B, and it seemed to sync well. I'm seeing my doc tomorrow to ask about adding Drug C to my current meds. For me personally Drug A is a very tricky drug, only helpful if I'm getting good sleep, eat and exercise well and limit the use to couple times a week."

Other software, he says, may only detect positive and negative words (such as "well" and "good" versus "tricky" and "limit"). DecisionEngine, on the other hand, would identify many more pieces of information, including the use and effectiveness of Drugs A and B combined; the dosage of Drug B; consideration for adopting Drug C; potential dissatisfaction with Drug A, depending on lifestyle choices such as "getting good sleep"; the commenter's use of three concurrent medications; and plans of visiting a health care professional.

These insights allow clients to take action, Nemirovsky says. If consumers are planning to switch drugs, for instance, a pharmaceutical firm may want to ensure that the consumers are using their products properly, and to find a means to address any issues.

Recently, Nemirovsky says, a pharmaceutical firm used DecisionEngine to determine if an allergy medication had improved the quality of life for a subgroup of patients. Analyzing specific issues associated with the subgroup, the firm discovered that the drug had an outsized positive impact, more so than several competing brands. The firm used the results in a regulatory submission—a critical stage in bringing any health

care product to market. "It's rare for the regulatory authorities to consider online patient reports as part of the regulatory approval process," Nemirovsky says.

## **Everyone's an expert**

In the late 2000s at MIT, Nemirovsky, who was an MIT Media Lab graduate student, and Quattoni, who was studying at the Computer Science and Artificial Intelligence Laboratory (CSAIL), came together with a lofty goal: Use big data to make everyone experts.

The plan was to combine machine learning with natural language processing to decode mountains of unstructured data and provide pertinent information, about anything, to anyone who wanted. "If you give people the right information, at the right time, anyone can be an expert," Nemirovsky says.

In building the software, they discovered that an important topic for most people on a daily basis is health care. "Patients go to the doctor with complex conditions, and sometimes they leave with less certainty they had before," Nemirovsky says. "Then they go online and say, 'What on Earth is going on? What do I do?'"

Focusing on the health care industry, they turned to MIT's Venture Mentoring Service, which helped them navigate various startup issues: fundraising, operations, marketing, legal issues, and other things.

"Things that sound obvious now, were not obvious to us at all," Nemirovsky says. "We were helped a lot by the VMS, especially as first-time entrepreneurs."

Soon after Nemirovsky graduated, he and Quattoni launched dMetrics in Boston, before relocating to Brooklyn. Over the years, the startup expanded from two to 16 employees—whose machine learning and

[natural language](#) processing research has been cited in more than 4,500 academic journals total—and earned four National Science Foundation grants to develop its technology.

Moving forward, dMetrics aims to bring its [software](#) to more sectors than [health care](#), politics, and consumer finance, with aims of empowering everyone with data. In that way, Nemirovsky says, the dMetrics mission hasn't changed much from its early MIT days: "It's our vision that we need to open means of expertise to everyone."

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

Provided by Massachusetts Institute of Technology

Citation: Detecting consumer decisions within messy data (2015, December 22) retrieved 11 May 2024 from <https://phys.org/news/2015-12-consumer-decisions-messy.html>

<p>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.</p>
------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------