

Bringing 'Common Sense' to Text Analytics

Rob Matheson, MIT News Office



Bringing “common sense” to artificial intelligence is one of the biggest challenges in computer science: It entails equipping computers with the shared knowledge that humans use to infer meaning, make connections and communicate, among other things.

Catherine Havasi dedicated more than a decade to such research, amassing an enormous knowledge base from around the Web. In 2010, she used that research as the technological foundation for Luminoso Technologies, a startup whose commercial software is helping bring common sense to text analytics.

Luminoso’s technology aims to quickly mine and analyze vast quantities of online text and— using a database of world knowledge— quickly identify opinions, patterns and underlying themes in the text. “It has this ‘backbone’ of common sense that allows our technology to spontaneously infer meaning” from text, Havasi says.

In a few short years, Luminoso has earned big-name clients, including Mars, BP and Scotts. The company’s tools have garnered praise in tech circles as valuable linguistic sources for, say, advertising and marketing agencies looking to gather more data about what people are saying about their products.

According to Havasi, the software’s “common sense” foundation helps it understand the quirks of human language— the allusions, jargon, cultural terms, shorthand and metaphors— that populate online chatter, helping it tease out overlooked connections and meanings in the text.

For instance, in an online review for a product, people may use cultural references, analogies and words with multiple meanings — such as “awesome” or “cool.” Humans, of course, can pick up on these linguistic tricks to decipher the overall meaning of the review, but software can’t. Luminoso tries to rectify that, says

Bringing 'Common Sense' to Text Analytics

Published on Scientific Computing (<http://www.scientificcomputing.com>)

Havasi, who is now an MIT research affiliate.

“It’s always been a tenet of linguistic philosophy that humans rely on unspoken assumptions about the world to understand one another, because we know a person is listening on the other end,” she says. “We try to help computers understand the world more like a person, as opposed [to] as a machine.”

Building ‘a backbone’ of common sense

The core of Luminoso’s technology dates back to 1999. As part of an Undergraduate Research Opportunities Program (UROP) project in the MIT Media Lab, Havasi worked under professor emeritus Marvin Minsky, who charged Havasi and two other grad students with a broad mission: giving computers common sense.

“It was one of those classic MIT moments, where they take a hearty artificial-intelligence problem and see what you can get done,” Havasi says.

The team— including Havasi’s collaborator, Pushpinder “Push” Singh ’98, PhD ’05— built a cloud-based crowdsourcing tool called the “Open Mind Common Sense” (OMCS) project. It collects information from Internet users who enter into a database various types of knowledge, such as word definitions and the relationships between words— insights such as, “The sun is hot.”

Networks called ConceptNet and AnalogySpace would then connect these concepts and infer new knowledge. Still freely accessible online, the project has since grown, having accumulated roughly 17 to 18 million data points from thousands of individuals.

Luminoso was essentially founded to provide a graphic interface and analytics engine for this project. But the technology has evolved drastically, into what Luminoso now calls its “conceptual engine” for analyzing messy text data.

Today, it sells largely to advertising and marketing agencies, as well as companies that want to know what people are saying about their products. If, for instance, an ad agency used the software to learn what people were saying about a product, the software would mine unstructured text— such as in news stories, research results or social media chatter— and shape all relevant words into a word cloud.

Within the cloud, the agency can search for, say, positive and negative sentiments— such as “love” and “hate”— and the words most associated with that adjective in the word cloud will turn a chosen color. So if the agency searched for “love,” all aspects of the product that customers love— such as its price or reliability— would appear in the cloud as, say, blue.

The technology has fleshed out some surprising, and useful, correlations between words. Once it revealed that users of certain toiletries spoke differently about brightly colored versions of the products. And it found that guests at an upscale

hotel were actually satisfied with paying more, feeling that they got more bang for their buck.

“What you want, as a client with a lot of data, is to be able to ask more difficult questions about your data and how these factors involved buying,” Havasi says. “We try to answer those questions.”

The ‘demo or die’ mentality

Although OMCS developed in a lab for more than a decade, the idea to commercialize the technology didn’t come until around 2010, when some of the Media Lab’s corporate sponsors began expressing interest in buying the technology.

To launch Luminoso, the founding team went through MIT’s Venture Mentoring Service, which connected them with lawyers and accountants, as well as business mentors, an advisory board and other entrepreneurial members of the MIT community.

Primarily, however, it was the Media Lab that helped Havasi with her entrepreneurial pursuits. For one thing, Media Lab students, she says, are constantly interacting with companies. Also, as a researcher in the Media Lab, she managed a dozen students and research projects that presented even more commercial possibilities.

“Then, the puzzle becomes: How do we take [the companies’] problems and apply them to our technology?’ The Media Lab understands companies and understands that you want to look to bring pieces that don’t necessarily go together, together,” she says.

Over the years in the Media Lab, Havasi says she unwittingly learned to test business assumptions, identify market needs, and network, among other things.

“The regular culture of the Media Lab is famously, ‘Demo or die.’ It’s the greatest place for doing product-market fit and product management, much more so than you know when you’re there,” she says. “I found a lot about starting companies at the lab without realizing it.”

Source URL (retrieved on 12/12/2013 - 6:03pm):

<http://www.scientificcomputing.com/news/2013/09/bringing-common-sense-text-analytics>