New research method predicts success, failure of D.C. restaurants

December 17 2014, by Greg Muraski

Using a database of 130,000 Yelp reviews of restaurants in Washington, D.C., two professors and a graduate student at the University of Maryland's Robert H. Smith School of Business have identified a method that allows software to "read" the content of those reviews—and predict which restaurants will close.

For the study, the researchers identified slightly more than 2,000 Washington, D.C., restaurants that were open as of December 2013. From various sources, they then identified roughly 450 that had closed from 2005 to 2014. To identify linguistic patterns that foretold closure, they paired restaurants according to such factors as price and cuisine type, and looked at how the descriptions varied.

It doesn't take a Ph.D. to know that there's a connection between a restaurants' Yelp rating and whether it will survive. But what Jorge Mejia, a UMD doctoral student; Shawn Mankad, an assistant professor; and Anandasivam Gopal, an associate professor have created is more powerful: Their computer-assisted text analysis proved more accurate at predicting restaurants' demise than ratings alone. And it is most powerful when used in combination with numerical ratings.

"The whole idea is that we are surrounded by all of this free, unstructured data," Mankad says—hundreds of thousands of words that would require armies of employees to read, let alone interpret. "We should be using that data."

The influence of online reviews is indisputable: More than 60 percent of Americans say that such reviews have high or medium level of influence over their buying decisions.

Other scholars have sought to take the emotional "temperature" of online reviews, by analyzing the proportion of positive versus negative words. This new approach goes deeper, examining constellations of words that were associated with restaurants' beating the long odds of their industry and remaining open.

For instance, restaurants for which reviewers used the words "food," "good," "place," "like," "order," "friend," "time," "great," "nice," and "service" tended to survive at unusually high rates. The Smith School

professors called the variable linked to those [words](link) "Quality_Overall," and it seemed to be the most potent signifier of general quality. "Constructing the variables, putting it into a predictive model—this is something that has never been done before," says Mankad.

They used one subset of data to uncover the relevant linguistic patterns and another subset to test the predictive power of their model. In that second group, the variables did predict, to a statistically significant degree, success or failure.

The working paper, "More Than Just Words: Using Latent Semantic Analysis in Online Reviews to Explain Restaurant Closures," grew out of Mejia's dissertation.

Mankad discussed the use of text data to improve business forecasting in October, at the Cornell Hospitality Research Summit. He also discussed the topic in a recent webinar sponsored by Hotel Management Magazine and the Wall Street Journal.

Provided by University of Maryland