Can machine learning deliver critical market insight on consumer needs faster and cheaper?

February 7 2019

Credit: CC0 Public Domain

Consumer brands have long used old-fashioned focus groups, interviews and surveys to best gauge consumer wants, desires and needs as part of processes that range from product development, to marketing and sales.

As machine learning and artificial intelligence (AI) have emerged, there is an increasing interest in the ability to harness these solutions to save time and money, and to yield more reliable consumer insights.

Machine learning can help to analyze user-generated content (UGC), which involves the collection of data from online reviews, social media, and blogs, that provide insights on consumer needs, preferences and attitudes.

Despite the potential for better information, marketers have raised concerns over the value of UGC data because the sheer scale and quality of UGC makes it difficult to process. While the data is accessible, identifying consumer insights requires human beings to analyze the data, which is hard to do at scale.

Two researchers from the Massachusetts Institute of Technology (MIT) decided to tackle this problem through research designed to examine the challenge of how to most efficiently use UGC to identify customer needs in ways that are more cost-efficient and accurate.

The study to be published in the February edition of the INFORMS journal *Marketing Science* is titled "Identifying Customer Needs from User-Generated Content," and is authored by Artem Timoshenko and John R. Hauser of MIT.

They find that machine learning can improve the process for identifying customer needs, while reducing research time substantially, helping consumer marketing brands avoid delays in bringing products to market.

"As more and more people turn to the digital marketplace to research products, share their opinions, and exchange product experiences, large amounts of UGC data is available quickly and at a low incremental cost to companies," said Timoshenko. "In many brand categories, UGC is

extensive.

For example, there are more than 300,000 reviews on health and personal care products on Amazon alone. If UGC can be mined for customer needs, it has the potential to identify customer needs better than direct customer interviews."

Other advantages of UGC data are that it is updated continuously, which enables companies to stay current with their understandings of customer needs. And unlike customer interviews, UGC data is available for research to return to further explore new insights.

To conduct their research, the study authors constructed and analyzed a custom data set which compares the customer needs for the oral-care category identified from direct interviews to the customer needs from Amazon reviews. The data set was constructed in a partnership with a marketing consulting firm to ensure the industry-standard quality of the interviews and insights.

The authors developed and evaluated a machine-learning hybrid approach to identify customer needs from UGC. First, they use machine learning to identify relevant content and remove redundancies. The processed data is then analyzed by human beings to formulate customer needs from selected content.

"In the end, we found that UGC does at least as well as traditional methods based on a representative set of customers," said Hauser. "We were able to process large amounts of data and narrow it to manageable samples for manual review. The manual review remains an important final part of the process, since professional analysts are best able to judge the context-dependent nature of customer needs."

More information: Artem Timoshenko et al. Identifying Customer

Needs from User-Generated Content, *Marketing Science* (2019). [DOI: 10.1287/mksc.2018.1123](https://doi.org)

Provided by Institute for Operations Research and the Management Sciences