Using machine learning and natural language processing to measure consumer reviews for product attribute insights

November 23 2021, by Marilyn Stone

Credit: Pixabay/CC0 Public Domain

Researchers from Western University, SUNY Buffalo State College, University of Cincinnati, and City University of Hong Kong published a

new paper in the *Journal of Marketing* that presents a methodological framework for managers to extract and monitor information related to products and their attributes from consumer reviews.

Understanding how concrete product attributes form higher-level benefits for consumers can benefit various corporate teams. Concrete, or "engineered attributes" refer to technical specifications and product features. For example, in the context of tablet computers, such attributes include RAM, CPU, weight, and screen resolution. Understanding how combinations of these lower-level attributes form higher-level benefits, or "meta-attributes," for consumers, such as Hardware and Connectivity, can provide managers with actionable insights. Sales teams need to understand the higher-level product benefits that drive consumer buying behavior. Product design teams must communicate with engineering and manufacturing to understand the relationships between the product's technical specifications and its perceived benefits. Engineering teams need to be able to estimate the trade-offs of technical subcomponents to build the product model that fulfills the more abstract benefits associated with the product's meta-attributes.

The traditional method of surveys can be time-consuming and may yield inconsistent results across different sampling periods. Thus, there remains a significant gap in theory and practice: How can the link between engineered attributes and meta-attributes be uncovered directly from consumer input to inform managerial decisions?

To fill this gap, the research team devised a methodological framework based on machine learning and natural language processing to obtain an embedded representation of product attributes. Specifically, embedded representation describes (represents) textual data such as individual product attributes using the words that surround such textual data (i.e., the contextual information) in consumer reviews. The representation is quantified using neural networks that enable mathematically

measurement of the degrees of similarity between various product attributes based on how they are described by consumers themselves (i.e., the contextual information), thus revealing similarities and differences in the attributes' usage by consumers.

From this embedded representation, the model then identifies multi-level clusters of product attributes that reflect the levels of abstract product benefits. "In other words," says Wang, "this new method algorithmically extracts consumers' own words in the reviews they write to quantify specific contexts that are expressed in relation to individual product attributes.

This then enables grouping the product attributes together based on their contextual similarities to uncover higher-level benefits that can influence consumer satisfaction or dissatisfaction with a product." The sentiments associated with these meta-attributes are used to evaluate objects of managerial interest, such as a product or brand, and then can go deeper to examine which engineered attributes primarily drive consumer sentiments in relation to the meta-attributes.

The research makes three main contributions. First, it provides a methodological framework for managers to extract and monitor information related to products and their attributes from consumer reviews. As He explains, "Because our framework exploits the contexts surrounding product attributes expressed in consumer reviews, managers can use it to directly monitor how meta-attributes evolve within brands and to compare brands within a product category to inform their product-related decisions. We provide validations that our hierarchical structure of meta-attributes adequately approximates consumers' underlying review-writing behaviors." Second, the research extends sentiment analysis of consumer reviews by demonstrating hierarchical sentiment analysis, which aggregates sentiment scores associated with individual attributes based on an attribute hierarchy.

Starting at the review level, sentiment scores can be aggregated upwards to yield insights for various units of analysis, such as SKU, product series, and brands. "Using hierarchical sentiment analysis, managers can go beyond relying on review ratings, which only describe products as a whole and cannot be accredited to specific product attributes. We demonstrate that this flexible approach to sentiment analysis can generate tailored dashboards and perceptual maps from consumer reviews that can inform managerial decisions," says Curry.

Third, the study uses consumer reviews of tablets to provide a practical demonstration of the method. In particular, it analyzes consumer sentiments about Hewlett-Packard and Toshiba to explore potential reasons why these brands ultimately discontinued their tablet product lines. Ryoo explains that "Using our attribute hierarchy, we evaluate their meta-attributes and then drill down to the level of engineered attributes to find that the limited number of apps available for HP's tablets and the thickness and weight of Toshiba's tablets were the main drivers of consumers' negative sentiments about the products.

We then analyze the meta-attributes of market-leading brands Samsung and Apple to explore potential drivers of their successes." Berger et al. note that "for data to be useful, researchers must be able to extract underlying insight—to measure, track, understand, and interpret the causes and consequences of market behavior." In this sense, this method is highly useful for developing marketing strategies because it provides valuable insights into the relationships between product attributes and consumer valuations.

Provided by American Marketing Association