U

Using Machine Learning to Extract Insights From Consumer Data

Hannah H. Chang
Singapore Management University, Singapore

Anirban Mukherjee
Cornell University, USA

INTRODUCTION

Advances in digital technology have led to the digitization of everyday activities of billions of people around the world, generating vast amounts of data on human behavior. From what people buy, to what information they search for, to how they navigate the social, digital, and physical world, human behavior can now be measured at a scale and level of precision that human history has not witnessed before. These developments have created unprecedented opportunities for those interested in understanding observable human behavior–social scientists, businesses, and policymakers—to (re)examine theoretical and substantive questions regarding people's behavior. Moreover, technology has led to the emergence of new forms of consumer marketplace—crowdfunding (whereby entrepreneurs obtaining funds from an anonymous online crowd; Mukherjee, Chang, & Chattopadhyay 2019) and crowdsourcing (whereby organizations gather new ideas and business solutions from an anonymous online crowd; Mukherjee, Xiao, Wang, & Contractor, 2018)—which not only details people's behavior in exchange of products and services but also led to new behavior.

Making sense of the vast amount of fine-grained data about consumer behavior, however, poses non-trivial challenges for marketing researchers and practitioners. In the past, behavioral data about consumers originated from sources such as point-of-purchase scanner data, customer attitude or satisfaction surveys, consumer purchase panels, and laboratory-based experiments. These traditional sources of consumer data are much smaller in scale, much more structured (e.g., in numbers-based data formats which can be directly analyzed), and measured to purpose, than new consumer data sources. Consequently, many of the methods used to analyze traditional customer data—such as conventional econometric and statistical methods—are not designed to deal with the breadth, precision, and scale of the new consumer data sources publicly available, which tend to be unstructured—written texts, images, audios, and videos—and require parsing and processing before data can be analyzed.

Fortunately, a parallel trend to the emergence of "big data" on consumer behavior has been the emergence of computational methods and analysis techniques needed to deal with these new sources of behavioral data—which tend to be more unstructured, of much larger scale, and noisier. Specifically, data on consumers come in four basic forms: (1) structured data, (e.g., number of likes on Facebook), (2) textual data (e.g., tweets on Twitter), (3) audial data (e.g., Spotify radio advertisements), and (4) visual data (e.g., photos on TripAdvisor). Consumer data can involve only one form, such as textual messages (e.g., tweets on Twitter) and visual images (e.g., Instagram photos). Consumer data can also involve more than one form. Video data, which are increasingly prevalent, combine a series of visual images (typically, 24 visual frames per second) and an audio track. Many publicly available sources of consumer relevant

DOI: 10.4018/978-1-7998-9220-5.ch107

data detailing people's consumption behavior involve multiple data elements. For example, consumer data from YouTube combines all four of these basic forms—audial and visual data in the video, textual data in the comments, and structured data in the number of views and likes. For each of these four data elements, new machine learning and big data methods enable us to simply and easily parse the data to uncover consumer insights if analysts are equipped with the right toolkit.

BACKGROUND

As both the availability of large-scale behavioral data and computational analysis methods are recent and emerging developments, many behavioral scientists and practitioners may be unaware or unfamiliar with (1) new sources of secondary data and types of data that are available to extracts insights about consumer behavior, and (2) new analysis techniques to study consumer behavior at scale. Therefore, motivated by these recent developments and opportunities, the main objective of this book chapter is to discuss computational methods (specifically, machine learning methods) for researchers and practitioners interested in addressing customer-relevant questions using new secondary data sources that are publicly available. This chapter offers a primer on the application of computational social science for understanding consumer data for both researchers and practitioners.

The rest of this chapter is organized as follows. First, types of unstructured data pertaining to consumer behavior—including the information that consumers are exposed to and their digital footprints in the modern marketplace—will be decomposed to their underlying data elements. Next, machine learning and computational techniques to parse and process unstructured customer data are described. Finally, potential directions for future research using consumer data are discussed.

TYPES OF CUSTOMER-RELEVANT DATA

Consumers today have unprecedented access to numerous types of information and media. New forms of information environments, often fueled by technology, have also emerged since the 2010s. For example, reward-based crowdfunding platforms like Kickstarter and Indiegogo post videos and text descriptions about new product innovations (Dhanani & Mukherjee, 2017; Allon & Babich, 2020). Debt-based crowdfunding platforms such as Lending Club and Funding Societies post text descriptions about business loans (Lee, Chang, & Mukherjee, 2020). As self-contained marketplaces, crowdfunding data include comprehensive descriptions of the factors that influence consumers and investors as well as detailed accounts of purchase and exchange behavior (Mukherjee et al., 2019). App-based ecosystems enable marketers to promote "green" initiatives as a part of their corporate social responsibility strategy (Merrill, Chang, Liang, Lan and Wong 2019). These platforms offer a wealth of detailed data to study emerging consumer behavior in new marketplaces. New voice assistants such as Amazon's Alexa and Apple's Siri allows consumers to search for product information through voice based commands, simplifying consumers' purchase journey, which led to changing customer behavior. This development also led to emerging forms of voice-assisted retail shopping that is often dubbed "voice commerce" (i.e., v-commerce), all of which offer a wealth of data to study emerging customer behavior in new marketplaces. Moreover, new digital channels help facilitate communication among businesses, consumers, investors, and other stakeholders.

Related Content