Chapter 19
Using Computational Text Analysis to Explore Open-Ended Survey Question Responses

Shalin Hai-Jew
Kansas State University, USA

ABSTRACT

To capture a broader range of data than close-ended questions (often defined and delimited by the survey instrument designer), open-ended questions, such as text-based elicitations (and file-upload options for still imagery, audio, video, and other contents) are becoming more common because of the wide availability of computational text analysis, both within online survey tools and in external software applications. These computational text analysis tools—some online, some offline—make it easier to capture reproducible insights with qualitative data. This chapter explores some analytical capabilities, in matrix queries, theme extraction (topic modeling), sentiment analysis, cluster analysis (concept mapping), network text structures, qualitative cross-tabulation analysis, manual coding to automated coding, linguistic analysis, psychometrics, stylometry, network analysis, and others, as applied to open-ended questions from online surveys (and combined with human close reading).

INTRODUCTION

The popularization of online surveys has meant that a wide range of different questions are ask-able, with the integration of still visuals, audio, video, web links, and other elements. Invisible or hidden questions enable the collection of additional information, such as time spent per question, devices used to access the survey, geographical information, and other data. File upload question types enable respondents to share imagery, audio, video, and other digital file types as a response. Integrations with online tools enable outreaches through social media for broader audiences through crowd-sourcing and commercial survey panels. Automation enables customizing survey experiences with uses of names, question an-

DOI: 10.4018/978-1-6684-6303-1.ch019

swers, piped text from a number of sources, expanded question elicitations (like through loop & merge techniques, and others), branching logic, and randomizers, among others. And many online research suites, designed as all-in-one shops, enable the automated analyses of text, quantitative data in cross-tabulation analyses, and other approaches.

Yet, in the midst of all these changes, a simple confluence of technological capabilities has suggested an even more fundamental change: the sophistication of computational text analysis (computer-aided text analysis) means that open-ended text-based survey question responses may be better harnessed and exploited for information than in the recent past. Computational text analysis enables the identification of a range of data patterns: matrix queries, theme extraction (topic modeling), sentiment analysis, cluster analysis (concept mapping), network text structures, qualitative cross-tabulation analysis, manual coding to automated coding, linguistic analysis, psychometrics, stylometry, network analysis, and others. These computational text analysis approaches harness quantitative, qualitative and mixed methods approaches, and all include "humans in the loop" for the analyses.

While some all-in-one online survey systems are expanding to built-in text analyses, the available tools look to be simplistic presently, with commercial software tools enabling more sophisticated text analysis. Those with the technology skills and statistical know-how stand to exploit the capabilities of open-ended survey questions and freeform respondent comments and insights. Going to "machine reading" (or "distant reading" through various forms of computational text analysis) does not remove the human from the loop. There is still the need for human "close reading" of the findings and of some of the original raw data. (In some cases, all of the original text may be read depending on the size of the text corpus.)

Technology Tools Used

The software tools highlighted in this work include Qualtrics®, NVivo 12 Plus, Linguistic Inquiry and Word Count (LIWC2015), and Network Overview, Discovery and Exploration for Excel (NodeXL).

REVIEW OF THE LITERATURE

The main strength of surveys is that they capture elicited information from human respondents, but that fact is also its main weakness. There is a wide body of literature that shows that people's responses to surveys may depend on social relationships, design features of how questions are presented and asked, the types of technologies used, and other factors, which "intervene" and "interfere" with respondents' offering their truest thinking. Besides these factors, the respondent himself/herself has limitations, in terms of built-in cognitive biases (confirmative bias, anchoring biases, priming effects, and others) and limited working memory. And yet, surveys are sometimes the only way to capture respondent experiences, preferences, imaginations, and opinions, even with the limitations of self-reportage.

Surveys are delivered in various ways. Surveys may be delivered in person or remotely, to respondents who are alone or in the company of others. They may be other-administered or self-administered. They may be delivered through various modalities: via telephone (Arnon & Reichel, Apr. 2009) or paper (postal or face-to-face) or computer, offline or online, and so on. There are some survey sequences that involve various mixes of the prior variables. Some classic Delphi survey methods began with face-to-face (F2F) meetings followed by distance-based interactions, for example.

Related Content