**International Journal of Social Media and Online Communities**
Volume 11 • Issue 2 • July-December 2019

# SMAGA:
## A Text Mining Framework to Study Culture and Cultural Differences

Yuan Xue, Pennsylvania State University, Berks, USA
Yilu Zhou, Fordham University, USA

### ABSTRACT

The study of national culture and cultural differences has immense value in today's business world. However, contemporary cross-cultural study instruments are subjective to many biases and intensive in terms of human labor. On the other hand, social media data offers tremendous opportunities for cross-cultural studies. However, it cannot be used directly as input for existing cross-cultural study applications in social science for being unstructured and noisy. In this work, the authors develop a text mining application called Social Media Associative Group Analysis (SMAGA). It effectively discovers information relevant to cultural values from social media data and represent such information using AGA verbal associations tuples. Experimental results show that the SMAGA framework can generate meaningful results in support of cross-cultural studies with much higher efficiency than traditional cross-cultural study instruments. This research underlines the emerging trend of developing text mining applications to automate cross-cultural and other types of social science studies.

### KEYWORDS

Associative Group Analysis, Cross-Culture Study, Data Analytics, Information Extraction, National Culture, Natural Language Processing, Social Media, Text Mining

### INTRODUCTION

The continuing trend of globalization has highlighted the importance of cross-cultural studies in many fields, including human resource management, marketing and strategic decision-making. Staff management within organizations involves managing cross-cultural issues since today's workforce consists of individuals with diverse ethnic backgrounds (Yayla-Küllü et al., 2015). Cross-cultural misunderstandings can be very costly in marketing as well; companies often have low product sales in foreign markets if their product commercials do not fit the local cultural values (Weiss, 1993). Cultural differences are also a key factor in decision-making at the strategic level, such as entry-mode choice (Kogut & Singh, 1988; Tihanyi et al., 2005) and merge and acquisition (M&A) (Stahl & Voigt, 2008; Weber et al., 1996). Additionally, cultural differences can also affect the outcomes of technology transfer and information exchange (Straub et al., 1997; Veiga et al., 2001). Thus, the study of culture and cultural differences is a high priority in the management science field.

DOI: 10.4018/IJSMOC.2019070102

Correspondingly, many instruments have been developed to study culture and cultural differences. A majority of these instruments are deductive, value-based surveys (VBS) (Bochner, 1994; Hofstede & Bond, 1984). Despite their popularity, there are unresolved issues regarding their reliability and validity (Schriesheim et al., 1993; Schwab, 1978), suitability for studying explorative topics such as culture (Hinkin, 1995) and sampling bias (Pinsonneault & Kraemer, 1993).

Associative Group Analysis (AGA) is an effective cross-cultural study instrument in the form of an open-ended survey. Given any cross-cultural study topic, an AGA survey can construct the culture-specific meaning of the concept using stimulus-associative keyword relations in linguistics. In comparison with the VBS approach, the AGA method has better uniformity, more objectivity and wider applicability in terms of its output. Therefore, AGA can produce valuable inputs for cross-cultural studies on diverse topics by generating a kaleidoscopic view of national culture and cultural differences across multiple dimensions. In spite of its potential as a powerful tool for studying culture and cultural differences, the AGA method is seldom used for large-scale cross-cultural studies focusing on multiple topics of interest because of the sampling challenges and high resource costs of conducting such surveys.

The use of social media data can lead to substantial improvements in the AGA methodology for social media text contents containing stimulus-associative keyword relations generated by a vast number of users. However, social media text contents do not exist in a structured format and contain considerable noise. These characteristics of social media data make it difficult to extract information relevant to cultural values from and to format such information for the use of cross-cultural analysis. In this study, we designed an innovative text mining framework called Social Media AGA (SMAGA). This application can effectively extract stimulus-associative relations from large amounts of social media text data and use the extracted semantic relation to represent the culture-related knowledge of social media users. The stimulus-associative relations generated in this way can be used as input for cross-cultural studies at both the macro and micro levels.

As far as we know, this study is one of the few that brings automation and improvements into cross-cultural studies using text mining techniques. We also hope that this study can encourage more intelligent applications that solve challenging problem in social science using technical knowledge from the text mining and data science fields.

## LITERATURE REVIEW

In this section, we first provide an overview of two mainstream survey instruments frequently used in contemporary cultural studies: VBS and AGA. This is followed by a discussion of their contributions and limitations in the context of cross-cultural studies. The prospects of using social media data analytics for cross-cultural studies are discussed in light of the limitations of VBS and AGA. We then present a major challenge in using social media data for cultural studies and propose our text mining-based framework, which effectively addresses this challenge.

### Value-Based Survey Instruments (VBS)

VBS is a popular tool for studying culture and cultural differences. VBS surveys are constructed using questions that measure the difference in the collective perceptions of different social groups on some universal human values. The social groups can be defined based on either national cultural backgrounds or organizational cultural backgrounds. Among all the VBS templates, the Rokeach Value Survey (RVS) (Rokeach, 1973; Rokeach & Ball-Rokeach, 1989), Schwarz Value Survey (SVS) (Schwartz, 1992, 1994) and Geert Hofstede's Value Survey Module (VSM) questionnaire (Hofstede, 1994; Hofstede et al., 2008; Hofstede & Minkov, 2013) are most often used in studying national cultural differences. RVS asks its participants to rank the importance of 18 terminal values, which reflect beliefs and lifetime goals, and 18 instrumental variables, which are preferable modes of behavior in achieving the terminal values. SVS contains 57 questions, which explore the differences