Probing Construct Validity in Data-driven Disaster Analysis

Wen-Ting Chung*, Yu-Ru Lin†
*Department of Psychology in Education, University of Pittsburgh
†School of Information Sciences, University of Pittsburgh
Email: wenting.chung@gmail.com, yurulin@pitt.edu

***Abstract*—In this position paper, we discuss the promise and peril in data-driven disaster analysis. We argue for the importance of being sensitive to the construct validity issue prevailed in many big data studies and propose a research strategy as a remedy for such issue. Our strategy comprises three steps: theory-driven set-up first, statistic assessment follows, and qualitative inquiry for further calibration. The goal is to translate activity signals captured from data to proper social or behavioral interpretation. We exemplify the use of the proposed research strategy through a study of risk perception following a disaster event, and discuss the strategy's potential and limitation.**

I. Introduction

The proliferation of social and human activity trace data provides unprecedented opportunity for further understanding and improving disaster response. Analysis from disaster-related activity traces could help elucidate human behaviors under realistic conditions of disaster situations, which have been difficult to capture with traditional social science methodologies such as surveys and laboratory experiments [1], [2]. However, most data science research stresses the importance of computational and statistical assessment while overlooks the meaning and interpretability of the behavioral measurement captured from the data [3]. Accordingly, data-driven analysis is subject to the threat of *construct validity* [4]. In this position paper, we argue the criticality of examining construct validity in data-driven analysis, and propose a research methodology to probe the construct validity issue.

Construct validity, originally derived from the concern of the truthfulness of whether a psychological test is adequately designed to measure a hypothesized human attribute or quality [5], has developed into an overarching concern of the validity of a research [4]. Construct validity involves how appropriate inferences can be legitimately made from the operationalizations in the study, and a major issue is how well the operational definition of a variable actually reflects the hypothesized concept. For example, while a study claims that a person holding a better sense of humor will lead a happier life, the specification of how "the sense of humor" and "happiness" are measured have decisive power of judging what the inference can be made. Shall we measure the numbers of jokes a person make each day to indicate the sense of humor, or shall we have a person read stories that are supposed to make people laugh and test if one does laugh or how loud one laughs? These different ways of measuring could lead to completely different recommendations to cultivate one's sense of humor and hence the study outcome is obscure.

The recent advance in big data and computational techniques allows researchers to extract or derive various activity descriptors (e.g., how many jokes a person makes each week in social media) and further to extract patterns of associations among activity descriptors in a relatively cheaper manner. The pattern extraction and assessment can start even before researchers decide *what to be measured*. Thus, the issue of construct validity tends to be overlooked in many big data studies.

We propose a *triangulation* research strategy as a remedy for probing the construct validity issues for data-driven analysis. The strategy comprises the following three steps.

1) Theory-driven set-up: The study agenda should be driven by (one or multiple) theoretical views of the concepts or constructs on which the effects and relationships are to be captured in data. This means the meaning of a construct should be defined or clarified before operationalization (i.e., creating measures for the construct), and the likely effects and relationships may be hypothesized based on the observation in similar context. A theory-driven research set-up is important to eliminate ambiguity in what meant to be captured, which is crucial for interpretability of the analysis outcome.
2) Statistic assessment: A proper operationalization relies on a well-defined construct and maps the construct to measurement that can be quantified in data. Many activity trace data contain rich activity descriptors that can be viewed as a "proxy" for measuring constructs as long as the operational definition and construct definition are consistent to some extent (e.g., the number of contacts and the breath of social network). Statistic assessment is expected to provide evidence to support the theoretical view(s) of constructs from the data.
3) Qualitative inquiry for calibration and new insights: Qualitative inquiry has multiple roles in further probing the construct validity. First, by looking into cases in data, it helps show that the effects or relationships of the measures are in reality consistent with the theoretical view. Second, it helps detect effects or relationships of the measures that are in fact inconsistent with the theoretical definition. Given the inconsistency, new theories and hypotheses may originate.

We exemplify the use of the proposed research strategy

using a study of risk perception following a disaster event [2].

II. Research Strategy: A Case Study

Study context: The 2015 Paris attacks occurred on November 13. In an immigrant society like the U.S., its immigration policy is sensitive to disruptive events that signal potential threat of any particular group of immigrants to its national security. It is thus indispensable to understand how this terrorist event would affect the U.S. public's perception of risk/danger toward specific immigrant groups, including Muslim and Latino, and how people from distinct political leanings may perceive risk differently. This study uses large-scale Twitter communications during the course of the event.

Step I: Theory driven research set-up: To establish a proper measurement, we need to identify the theoretical views of risk perception relevant to our research context. Our conceptualization of *group risk perception* was grounded in social and cultural psychology studies – in particular, *moral dyad theory* [6], *moral foundation theory* [7], and *social identity theory* [8]. These theories offer probable relationships to be looked for in data. For example, social identity theory [8] suggests that people categorize social groups as ingroup or outgroup, depending on how they identify themselves as a belonged group member. People tend to favor their identified ingorup over outgroup, risk perception is hence captured by how distinct groups of people within a society identify the *Self* and the *Other*, as the other is often more likely to be perceived as a dangerous threat and potential harm [9].

Step II: Measurement computing and statistical assessment: We operationalized the construct of group risk perception by leveraging moral foundation lexicon [10]. Using this lexicon, we quantitatively measured risk of different groups through capturing their tweet expressions indicating a group/issue was morally judged or perceived as a risk issue. The operationalizaiton allowed us to generate risk profiles (as shwon in Fig. 1) for different groups and assess the statistic significance of the difference between groups as well as the change in each group following the event.

Step III: Qualitative inquiry for calibration: We conducted qualitative discourse analysis on a randomly sampled of 17,913 tweets. The analysis first examined whether those tweet contents identified as engaging in certain type of moral judgement through the lexicon approach were meaningful, and furthermore, differentiated whether an immigrant group was perceived as risk or being at risk, which was not able to be captured by Step II. Our results showed that there appeared a spectrum of how the tweeters differentiated Muslim group, Islam religion, ISIS, and terrorism/terrorist from one another – on one end, Muslim, Islam, and ISIS were discussed as if they equated to one another and were the terror itself; on the other, it was argued that Muslims and Islam did not equate to ISIS, and ISIS was not real Muslim but the terrorists claiming a Islam religious root. Moreover, the analysis suggested that group boundary may shift due to members' agreement or disagreement of risk perception, which potentially leads to further development of existing theories.

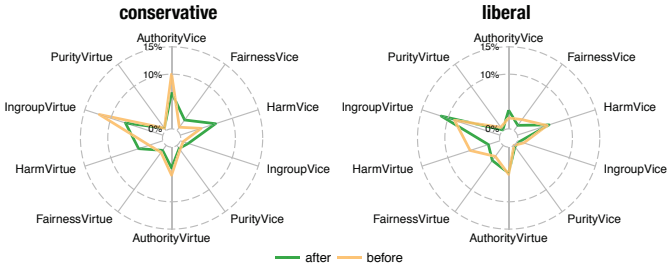

Fig. 1. Users' moral dimensions exhibited significant differences toward different immigrant groups (Latino, Muslim, etc.) before and after the attacks, and across leanings.

III. Discussion and Future Work

This paper outlines our approach to probe the construct validity issue when conducting data-driven analysis for understanding human response during disasters. Using a case study, we illustrate how the proposed research strategy can avoid the misinterpretation of study outcome. One limitation of this approach is that the sequence of the steps may restrict the extent of to which a study can explore when prior theories are insufficient or inadequate. We envision the qualitative inquiry is the key to establish new theories and plan to develop research guideline to closing the loop.

Acknowledgment

This work is part of the research supported from NSF grant #1423697, #1634944 and the CRDF at the University of Pittsburgh. Any opinions, findings, and conclusions or recommendations expressed in this material do not necessarily reflect the views of the funding sources.

References

[1] W. X. and L. Y.-R., "Sensing distress following a terrorist event," in *In 2016 International Conference on Social Computing, Behavioral-Cultural Modeling & Prediction and Behavior Representation in Modeling and Simulation (SBP-BRiMS 2016)*, 2016.

[2] W.-T. Chung, K. Wei, Y.-R. Lin, and X. Wen, "The Dynamics of Group Risk Perception in the US After Paris Attacks," in *Proc. of the 8th International Conference on Social Informatics (SocInfo 2016)*, 2016.

[3] D. Boyd and K. Crawford, "Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon," *Information, communication & society*, vol. 15, no. 5, pp. 662–679, 2012.

[4] S. Messick, "Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning." *American psychologist*, vol. 50, no. 9, p. 741, 1995.

[5] L. J. Cronbach and P. E. Meehl, "Construct validity in psychological tests." *Psychological bulletin*, vol. 52, no. 4, p. 281, 1955.

[6] K. Gray, L. Young, and A. Waytz, "Mind perception is the essence of morality," *Psychological Inquiry*, vol. 23, no. 2, pp. 101–124, 2012.

[7] J. Graham, J. Haidt, S. Koleva, M. Motyl, R. Iyer, S. P. Wojcik, and P. H. Ditto, "Moral foundations theory: The pragmatic validity of moral pluralism," *Advances in Experimental Social Psychology, Forthcoming*, 2012.

[8] H. Tajfel, *Social identity and intergroup relations*. Cambridge University Press, 2010.

[9] D. Lupton, "Risk as moral danger: the social and political functions of risk discourse in public health," *The sociology of health and illness*, 2001.

[10] J. Graham, J. Haidt, and B. A. Nosek, "Liberals and conservatives rely on different sets of moral foundations." *Journal of personality and social psychology*, vol. 96, no. 5, p. 1029, 2009.