Aug 10th, 12:00 AM

# Machine Learning and Grounded Theory: New Opportunities for Mixed-Design Research

Jang Bahadur Singh

M Vimal Kumar

Rajesh Chandwani

Biju Varkkey

# Machine Learning and Grounded Theory: New Opportunities for Mixed-Design Research

*Emergent Research Forum (ERF)*

**Jang Bahadur Singh**
Indian Institute of Management
Tiruchirappalli
jbs@iimtrichy.ac.in

**M Vimalkumar**
Indian Institute of Management
Tiruchirappalli
Vimalkumar.f15005@iimtrichy.ac.in

**Rajesh Chandwani**
Indian Institute of Management
Ahmedabad
rajeshc@iima.ac.in

**Biju Varkkey**
Indian Institute of Management
Ahmedabad
bvarkkey@iima.ac.in

## Abstract

Grounded theory (GT) can be simply understood as *"the discovery of emerging patterns in data"*[1]. Classic GT (or Glaserian GT[2]) is more specifically known for providing a general methodology for the development of theory inductively using any data (Holton & Walsh, 2016). Since its introduction (Glaser & Strauss, 1967), GT had a significant impact on theory building through qualitative inquiry in various disciplines, for example in Information systems, seminal work of IT and organizational change (Orlikowski, 1993) and cross-cultural collaboration in virtual teams (Levina & Vaast, 2008). Though GT's approach accommodates any type of data with its dictum, "All is Data." But GT studies using quantitative data are limited; even in mixed design studies, its role is limited to "basic descriptive statistics, correlation, and cluster analysis" (Walsh et al., 2015).

In a digitized world, organizations record/collect/log their activities in the form of digital "traces data[3]." Since such data is not collected with any particular research objective, this data is simply amassed in the organization's information system (Howison, Wiggins, & Crowston, 2011). From a social science and organization research perspective, such a new source of data can provide many research opportunities (Berente, Seidel, & Safadi, 2019). Particularly, with the assistance of machine learning (ML) algorithms, which are derived at the intersection of statistics and computer science, we can discover robust patterns in traces data. Such patterns could provide powerful insight related to social life (Puranam, Shrestha, He, & von Krogh, 2018). However, since ML-based techniques are primarily geared towards the task of prediction, getting "Y hat", not "beta" (Mullainathan & Spiess, 2017), such patterns or *empirically regularities* are not considered as "theory" among scholarly outlets (Agarwal & Dhar, 2014; Walsh et al., 2015). Also, when the

[1] This was a response to the question of "What Grounded theory is?" by Barney Glaser at the GT panel symposium of 2013 annual meeting of AOM in Orlando, Florida USA. Subsequently, conversation among panelist appeared as an article in Organizational Research Methods. For details please see Walsh et al, (2015).

[2] Barney Glaser and Anselm Strauss are the founders of GT with their seminal book Glaser and Strauss (1967), however later they fell into disagreement and developed GT further independently. Glaser's approach is known as classic GT or Glaserian GT. For details on their disagreements, please see (Heath & Cowley, 2004).

[3] Traces data is nothing but the log of records of every action/transactions and it is collected due to underlying use of information technology systems (if any). More info at: https://www.igi-global.com/dictionary/research-methodological-foundations-transaction-log/30305

number of variables are large (more likely in traces data), the likelihood of having highly correlated variables is significant, which in turn can lead to the substitution of variables in iterations of ML models during training, without affecting the accuracy of the model. The very idea of the ability of ML techniques to identify hidden generative structures within data raises the question about the stability of these patterns. Thus, Mullainathan and Spiess (2017, p-98) stress that future research in machine learning would be to make sense of "some structures" that ML models have found "without making strong assumptions about the underlying true world."

In recent years, there is significant attention being paid towards incorporating ML methods in management and organization research. For instance, Puranam et al. (2018) present well-argued guidelines for "algorithmic induction" through ML methods that could be used as a tool in interpretative and comparative case analysis. Choudhury et al (2018) present steps to generate robust patterns from data through ML which later can be tested with traditional econometric methods. While these studies bring ML techniques into mainstream management and organization research, its application is still portrayed as an exploratory tool. However, there is an opportunity to combine ML methods to grounded theory method (GTM), which would strengthen the researcher's ability to develop "stronger hybrid theory" (Muller, Guha, Baumer, Mimno, & Shami, 2016). As we note earlier GTM is open to all types of data and analysis techniques. GTM emphasizes that quantitative/statistical analysis methods are "merely techniques for arriving a type of fact, it is still up to the researcher to discover and analyze the theoretical relevance of these facts" (Glaser and Strauss, 1967, p-201).

However, to the best of our knowledge, empirical studies have not yet systematically attempted to blend ML with GT methods in their research design which could be helpful in better understanding of the research phenomenon. In our ongoing research project, we present our early experience of blending machine learning methods with qualitative data based on the framework of mixed-design grounded theory method (Walsh, 2015). In this ongoing research, we discuss emerging ML-driven mixed-design approach. We started with a large volume of traces data related to the employees of a large organization over the years; this slice of data was analyzed using supervised machine learning techniques[4] with a broad area of inquiry of fair promotions (presence or absence of Gender discrimination in promotions or Glass Ceiling). The pattern that emerged through ML-based techniques were further analyzed with Glaser's (Glaser, 2008) elaboration analysis of theoretical ordering. Such analysis helped in finding the core concepts (patterns) that could largely explain the variation in the data. These emerging patterns were further used for the theoretical sampling of qualitative inquiry. It provided another slice of data that helped in conceptualization through constant comparative analysis. We emphasize here that mixing ML techniques and qualitative methods with principles of GT, one could avoid the suspicion of correlated covariates, spurious factors and anticipated consequences (Walsh, 2015). ML method's Achilles heel of not knowing conclusively the covariates responsible for variation in data (Mullainathan and Spiess, 2017) could be mitigated with the mixed-design GT framework. In this paper, we share our initial results and emphasize that such a mixed methodological framework yielded relationships that otherwise would have been difficult to preconceive or time-consuming to observe.

**Keywords**

Machine Learning, Grounded Theory, Mixed-Design Research, Methodology.

# REFERENCES

Agarwal, R., & Dhar, V. (2014). Big Data, Data Science, and Analytics: The Opportunity and Challenge for IS Research. Information Systems Research, 25(3), 443–448.

[4] Machine learning algorithms are broadly categorised into supervised and unsupervised learning algorithms. They both seek to discover patterns in data automatically, while supersized ML algorithms need paired input-output data to find the pattern among them. Unsupervised algorithms summarize the input data without any output or target variable. There are various supervised ML techniques (such as decision trees, Support vector machine, neural networks etc.) with varying degree of interpretability. While decision trees are most interpretable, neural networks are like black box. In this study, we have chosen decision tree algorithms due easy interpretabilities of their outcome which can easily be mixed with GT approach. For details various machine leaning algorithms, please see Choudhury et al (2019) and Puranam et al, (2018).

Berente, N., Seidel, S., & Safadi, H. (2019). Research Commentary—Data-Driven Computationally Intensive Theory Development. *Information Systems Research*, *30*(1), 50–64.

Choudhury, P., Ryan, A., & Endres, M. G. (2018). Machine Learning Complementarities with Regression Analysis: Selection, Evaluation, and Interpretation of Machine Learning Methods for Exploratory Data Analysis and Pattern Detection. *Harvard Business School Working Paper*, *19–032*.

Glaser, B. (2008). Doing Quantitative Grounded Theory. Sociological Press.

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory : strategies for qualitative research*, Routledge.

Heath, H., & Cowley, S. (2004). Developing a grounded theory approach: a comparison of Glaser and Strauss. *International Journal of Nursing Studies*, *41*(2), 141–150.

Holton, J., & Walsh, I. (2016). *Classic grounded theory: Applications with qualitative and quantitative data*, SAGE Publications.

Howison, J., Wiggins, A., & Crowston, K. (2011). Validity Issues in the Use of Social Network Analysis with Digital Trace Data. *Journal of the Association for Information Systems*, *12*(12), 767–797.

Levina, & Vaast. (2008). Innovating or doing as Told? Status Differences and Overlapping Boundaries in Offshore Collaboration. *MIS Quarterly*, *32*(2), 307.

Mullainathan, S., & Spiess, J. (2017). Machine Learning: An Applied Econometric Approach. *Journal of Economic Perspectives*, *31*(2), 87–106.

Muller, M., Guha, S., Baumer, E. P. S., Mimno, D., & Shami, N. S. (2016). Machine Learning and Grounded Theory Method: Convergence, Divergence, and Combination. In *Proceedings of the 19th International Conference on Supporting Group Work - GROUP '16* (pp. 3–8). New York, New York, USA: ACM Press.

Orlikowski. (1993). CASE Tools as Organizational Change: Investigating Incremental and Radical Changes in Systems Development. *MIS Quarterly*, *17*(3), 309.

Puranam, P., Shrestha, Y. R., He, V. F., & von Krogh, G. (2018). Algorithmic Induction Through Machine Learning: Opportunities for Management and Organization Research. *SSRN Electronic Journal*.

Walsh, I. (2015). Using quantitative data in mixed-design grounded theory studies: an enhanced path to formal grounded theory in information systems. *European Journal of Information Systems*, *24*(5), 531–557.

Walsh, I., Holton, J. A., Bailyn, L., Fernandez, W., Levina, N., & Glaser, B. (2015). What Grounded Theory Is…A Critically Reflective Conversation Among Scholars. *Organizational Research Methods*, *18*(4), 581–599.