

Literature Review on the Cross-Section of Data Science, Machine Learning, and Anthropology

By Mayur Chavan, Sabastion Halverson

Abstract—At first glance, the fields of machine learning and cultural anthropology seem to have little overlap. Machine learning is about big data, predictions, and the generalized picture, while cultural anthropology focuses on the every-day context of people's lives from a very qualitative standpoint. However, we show that there lays a fertile field of potential that is yet to be cultivated into a full-grown garden of discovery between these two fields, if you can get past the methodological differences of big data and the ethnography.

Index Terms—Machine learning, data analysis, ethnography

I. INTRODUCTION

On the surface, the fields of cultural anthropology and machine learning could not seem more different. The ethnography, the core research philosophy of anthropology, is a personal recounting of an anthropologist's experience participating and observing individuals of a culture. It is in essence a deep dive into the context of culture on the most base level. Contrast this with data science and machine learning, disciplines based on generalization, prediction, and the "big data" needed to make the former run, which often has issues dealing with the context of the individual. Indeed, one could say that they are nearly incompatible, however, we argue that they instead cover each others' blind spots and can work together in ways the field has barely explored. We will also admit that our lack of experience in machine learning itself meant that we may have missed some papers on the implementation of machine learning techniques as we lack context on to what relevant search terms to use, and how to find out if an article would be relevant to our research even if we did.

II. METHOD

From the very start of our research, we found that this intersection of anthropology and machine learning is one not yet widely studied. Multiple library search engines, such as Parkside's, returned none or one relevant result. In all of the journals we looked in, no more than five articles could be found that were completely relevant to the topic in any single one. Given the condition of a very sparse field split between many different databases, we agreed to record the search parameters and engine used to find each relevant article, which are presented in the Results section.

For the selection of search terms, "machine learning" in anthropology journals was used, while "machine learning anthropology" was a better fit for databases that weren't specialized in anthropology. In our search we found quite a few articles that were near to our topic, but not exactly relevant. We chose not to include papers on sub-fields on anthropology

such as forensic, biological, or evolutionary anthropology, as those deal more with the physical aspects of humans rather than culture. Additionally, archaeology or linguistics were not included if they did not pertain to the study of culture in particular. Finally, anthropology papers that talked about the culture of machine learning without discussing machine learning itself were not included, as interesting as an ethnography of scientists is.

III. RESULTS

In this section, each of the articles we found are listed, along with the associated search term and database they were found in. An Overview of the article's purpose and relevancy to the topic is included.

Machine Learning Applications in Anthropology: Automated Discovery over Kinship Structures [1] - *Search Terms*: machine learning anthropology, *Found With*: Parkside Library, *Overview*: A common problem in anthropological field work is generalizing rules governing social interactions and relations (particularly kinship) from a series of examples. One class of machine learning algorithms is particularly well-suited to this task: inductive logic programming systems, as exemplified by FOIL. This is one of the first examples of machine learning in anthropology work.

The Stakes of Uncertainty: Developing and Integrating Machine Learning in Clinical Care [2] - *Search Terms*: machine learning, *Found With*: AnthroSource, *Overview*: Dr. Elish performed an ethnography on the development of a medical machine learning application, dealing with how both staff and users deal with "trusting" machine learning's predictions. Dr. Elish goes much deeper into the philosophy of data science itself, a bit outside the purview of our research.

Hybrid Methodology: Combining Ethnography, Cognitive Science, and Machine Learning to Inform the Development of Context-Aware Personal Computing and Assistive Technology [3] - *Search Terms*: machine learning, *Found In*: AnthroSource, *Overview*: A conceptual deep dive in combining data science, cognitive science, and the ethnography to create machine learning applications that can consider different kinds of context.

Anthropology by Data Science [4] - *Search Terms*: machine learning, *Found In*: AnthroSource, *Overview*: This essay explores how to broaden the scope of what constitutes anthropological and ethnographic research by cross-fertilizing with data science. The author argues that anthropologists should adopt an anthropology by data science perspective through incorporating machine-learning and other data science techniques into anthropological research.

CARE AND SCALE: Decorrelative Ethics in Algorithmic Recommendation [5] - *Search Terms*: machine learning, *Found In*: AnthroSource, *Overview*: A look into the ethics and values lost in the conflict between the context and individuality lost when scaling a service or algorithm to cover larger numbers of people. This looks at the context of people making machine learning algorithms and the values they try to preserve. The researcher challenges conventional anthropological thought by borrowing from machine learning concepts.

Big Data, Actually: Examining Systematic Messaging in 188 Romantic Comedies Using Unsupervised Machine Learning [6] - *Search Terms*: machine learning culture, *Found In*: APA PsysNet, *Overview*: Popular films within the romantic comedy genre set social expectations for romantic relationships, sexual activity, and gender roles, especially for young audiences. Nevertheless, academic research on the genre is scarce and mostly limited to manual content analyses conducted on relatively small samples of films. This study utilizes machine learning techniques to analyze the top-188 grossing "rom-coms" for common themes and traits.

A machine learning model of cultural change: Role of prosociality, political attitudes, and Protestant work ethic. [7] - *Search Terms*: machine learning culture, *Found In*: APA PsysNet, *Overview*: The researchers utilized machine learning to analyze a large amount of value surveys conducted over periods of cultural change to identify if certain traits were markers of this cultural change. This research illustrates that the abductive approach of machine learning, which focuses on the most likely explanations for an outcome, can help generate novel insights.

Anthropology and the AI-Turn in Global Governance [8] - *Search Terms*: machine learning anthropology, *Found In*: American Journal of International Law, *Overview*: This article serves to advocate the ethnography to determine the effect AI would have and does have on the every-day person who has to be affected by it, especially if it is integrated into law, and how would these uncovered patterns be realistically acted upon by humans?

IV. ANALYSIS

A. Societal Impact

The fields of both machine learning and cultural anthropology have a lot to offer society. Machine learning's predictions can assist in many ways, such as assisting doctors' in assessing risk of certain diseases [2] as well as the quality of life with helpful music recommendations [5] and numerous other applications. The ethnography of cultural anthropology serves as powerful tool to analyze the every-day context of people that can be missed by making generalizations over large groups, as well as making insights that can be missed by only looking at the numbers.

Several articles found during research discuss a variety of ways these two fields already intersect and will intersect more in the future. For one, anthropologists can assist in the roll-out of machine learning by observing why people don't trust a new technology [2] by assessing every person

in the pipelines, from developer, to doctor, to end point user. When this is applied to assist people in law, anthropologists can ask the questions they always do with new technology: How is it actually used by every-day people? Who controls this technology, and can/do they use this control to influence others? In the case of law, how does a system such ingrained in culture get solidified into code? These questions are pertinent as ever and anthropologists are the ones who seek the answers [8]. In another case, anthropologists and data scientists can work together to bridge the gap between the specific and the general, to find a way to scale machine learning for everyone while trying not to lose the many kinds of context that exists for individuals [5].

B. Methodology

For a few of these articles, combining the methodologies of cultural anthropology, the ethnography, and data science, machine learning, is a major goal of the piece. One researcher argues that anthropologists must introduce more data science and machine learning techniques holistically into their research to use technology to its greatest potential [4].

Another research group, including personnel from Facebook, goes even deeper to suggest a combination of anthropology, data science, and cognitive science, to create an ethnography that can be used in machine learning applications, solving the problem of context [3]. Machine learning has historically been fairly poor at understanding all the different types of context a situation may have, be it cultural, spatial, temporal, or any other. The ethnography, on the other hand, is the epitome of context, it seeks to understand and give meaning to even the most mundane of everyday activities.

C. Practical Applications

In our research, we found many fascinating articles going into the deep ethical, methodological, and intersecting ways in which this cross-pollination of fields can occur, however, there were hardly any real-world studies conducted that utilized both machine learning and anthropology methods, though the few we did find were promising and will serve as inspiration for our own project. The first recorded use of machine learning we found was in '96 with the application of machine learning to determine general rules in complex kinship systems for field researchers utilizing inductive logic programming systems, which excels in relational data [1]. This program exceeded its traditionally engineered counterpart in many areas, but not all.

Another rather amusing, but still relevant, article was on using machine learning to analyze a number of different culturally prevalent themes in the top-188 grossing romantic comedies [6]. While conducted not by an anthropologist, but a communications researcher, it is still analyzing culture and utilizing machine learning, which is a major example of how machine learning can help parse through a massive amount of data that would be either very time consuming or difficult for a human to accomplish. This utilization of machine learning

to help researching gleam novel insights is a very likely applicable use for machine learning in anthropology research.

In the third and final example we could find of machine learning application with cultural anthropology, where researchers utilized a large amount of surveys that polled respondents of their values [7]. These surveys were from a wide range from places over the globe and over a large period of time. The researchers were able to use this to analyze the change of importance in cultural values over time and find correlation between certain values changing to signify overall cultural change.

V. WHEN THE MACHINE MEETS THE EXPERT: AN ETHNOGRAPHY OF DEVELOPING AI FOR HIRING

The introduction of machine learning (ML) in organizations comes with the claim that algorithms will produce insights superior to those of experts by discovering the “truth” from data. Such a claim gives rise to a tension between the need to produce knowledge independent of domain experts and the need to remain relevant to the domain the system serves. This two-year ethnographic study focuses on how developers managed this tension when building an ML system to support the process of hiring job candidates at a large international organization. Despite the initial goal of getting domain experts “out the loop,” we found that developers and experts arrived at a new hybrid practice that relied on a combination of ML and domain expertise. We explain this outcome as resulting from a process of mutual learning in which deep engagement with the technology triggered actors to reflect on how they produced knowledge. These reflections prompted the developers to iterate between excluding domain expertise from the ML system and including it. Contrary to common views that imply an opposition between ML and domain expertise, our study foregrounds their interdependence and as such shows the dialectic nature of developing ML. We discuss the theoretical implications of these findings for the literature on information technologies and knowledge work, information system development and implementation, and human–ML hybrids.

VI. PERSONALIZED QUANTIFICATION OF FACIAL NORMALITY: A MACHINE LEARNING APPROACH

What is a normal face? A fundamental task for the facial reconstructive surgeon is to answer that question as it pertains to any given individual. Accordingly, it would be important to be able to place the facial appearance of a patient with congenital or acquired deformity numerically along their own continuum of normality, and to measure any surgical changes against such a personalized benchmark. This has not previously been possible. We have solved this problem by designing a computerized model that produces realistic, normalized versions of any given facial image, and objectively measures the perceptual distance between the raw and normalized facial image pair. The model is able to faithfully predict human scoring of facial normality. We believe this work represents a paradigm shift in the assessment of the human face, holding great promise for

development as an objective tool for surgical planning, patient education, and as a means for clinical outcome measurement.

VII. CONCLUSION

From our research, we have found that there has been a large amount of research setting the stage and laying down methodology in recent years, and we believe our research will be able to strike this while the iron is hot. Though there are not many examples, those that do exist show that data science and anthropological data can be combined to reach conclusions neither could have reached by itself. This relative lack of current applications simply means we have more reign to find an area of application ourselves without fear of redundancy. We believe that the examples that exist give us the framework we need, examples to draw on, and the data we’ll need to use in machine learning.

REFERENCES

- [1] S. J. Cunningham, “Machine learning applications in anthropology: Automated discovery over kinship structures,” *Computers and the Humanities*, vol. 30, no. 6, pp. 401–406, 1996. [Online]. Available: <http://www.jstor.org/stable/30204664>
- [2] M. C. ELISH, “The stakes of uncertainty: Developing and integrating machine learning in clinical care,” *Ethnographic Praxis in Industry Conference Proceedings*, vol. 2018, no. 1, pp. 364–380, 2018. [Online]. Available: <https://anthrosource.onlinelibrary.wiley.com/doi/abs/10.1111/1559-8918.2018.01213>
- [3] M. CURY, E. WHITWORTH, S. BARFORT, S. BOCHEREAU, J. BROWDER, T. R. JONKER, K. S. KIM, M. KRENCHER, M. RAMSEY-ELLIOT, F. SCHÜÜR, D. ZAX, and J. ZHANG, “Hybrid methodology: Combining ethnography, cognitive science, and machine learning to inform the development of context-aware personal computing and assistive technology,” *Ethnographic Praxis in Industry Conference Proceedings*, vol. 2019, no. 1, pp. 254–281, 2019. [Online]. Available: <https://anthrosource.onlinelibrary.wiley.com/doi/abs/10.1111/1559-8918.2019.01284>
- [4] S. Paff, “Anthropology by data science,” *Annals of Anthropological Practice*. [Online]. Available: <https://anthrosource.onlinelibrary.wiley.com/doi/abs/10.1111/napa.12169>
- [5] N. SEAVER, “Care and scale: Decorrelative ethics in algorithmic recommendation,” *Cultural Anthropology*, vol. 36, no. 3, pp. 509–537, 2021. [Online]. Available: <https://anthrosource.onlinelibrary.wiley.com/doi/abs/10.14506/ca36.3.11>
- [6] M. Moore and Y. Ophir, “Big data, actually: Examining systematic messaging in 188 romantic comedies using unsupervised machine learning,” *Psychology of Popular Media*, 04 2021.
- [7] A. Sheetal and K. Savani, “A machine learning model of cultural change: Role of prosociality, political attitudes, and protestant work ethic,” *American Psychologist*, 09 2021.
- [8] M. Sapignoli, “Anthropology and the ai-turn in global governance,” *AJIL Unbound*, vol. 115, p. 294–298, 2021.
- [9] E. van den Broek, “When the machine meets the expert: An ethnography of developing ai for hiring,” *Mis Quarterly*, vol. 30, no. 6, pp. 401–406, 2021.
- [10] M. Kukla-Bartoszek, “On the use of machine learning algorithms in forensic anthropology,” *Computers and the Humanities*, vol. 30, no. 6, pp. 401–406, 2020.
- [11]
- [12] M. Kukla-Bartoszek, “Searching for improvements in predicting human eye colour from dna,” *Computers and the Humanities*, vol. 30, no. 6, pp. 401–406, 2021. [Online]. Available:

ACKNOWLEDGMENT

Thank you to Dr. Kathleen Gillogly for help in topic refinement, data gathering, and expert advice.