Julia Bullard & Muhammad Abdul-Mageed
University of British Columbia, Vancouver, BC, Canada

Machine Knowledge: Information Studies and Artificial Intelligence in Dialogue (Paper)

Abstract:

In this piece, we examine the parallels and contrasts between knowledge organization and data science with a focus on the subfields of AI and machine learning in particular. Our analysis aims to both highlight and aid making connections between information studies and artificial intelligence through the discussion of shared practical and theoretical considerations and concerns around simplification and transparency.

1. Introduction

Data is the center of a great deal of excitement and trepidation in industrial, popular, and academic spaces. Computational approaches that exploit large-sized datasets are becoming increasingly mainstream in many fields as they offer new possibilities for analysis and making discoveries. However, this promise comes alongside the realization that we do not yet fully understand the range and scope of potential, unintended consequences deploying automated methods can result in. Of particular concern are methods that can encode historical discrimination into the methodological tools of the state (Angwin et al., 2016) and of researchers (Blodgett & O'Connor, 2017). With each advance in the power of data science to automate and scale decision-making we are likely to identify more 'blind spots' (Crawford & Calo, 2016) in how algorithms can misrepresent vulnerable social groups, thus creating unjust outcomes and public scandals.

Reflection within the artificial intelligence (AI) and machine learning communities (e.g., Barocas & Boyd, 2017) has drawn attention to a number of underlying problems that produce such outcomes and scandals: the complexity of the social world, value conflicts, and a lack of communication about ethics in trade-offs. However, the enthusiasm for the potential of AI approaches and the speed at which they might be developed and deployed are out of sync with the speed of the critical scholarship in response, such that much of the conversation around these issues is taking place among preprint article repositories (Wang & Kosinski, 2017), the popular press (Levin, 2017a, 2017b), and blog platforms (Agüera y Arcas, Todorov, & Mitchell, 2018), as in the case of the recent debate around the possibility of AI approaches to recognize sexual orientation from portraits. As a complementary approach to responding to new cases as they arise, we suggest looking back at the history of discussion around similar issues in knowledge organization.

2. AI and Knowledge Organization

We recognize numerous parallels between the challenges facing AI and machine learning scholars and long-standing debates in knowledge organization. For example, contemporary findings on how natural language processing might discriminate against African-American voices (Blodgett & O'Connor, 2017) parallel cornerstone discussions on how classification systems make some people and experiences invisible (Star & Bowker, 2007). These similarities are not by chance; rather, they result from the shared aim of machine learning and knowledge organization: to address the messiness of human expression, language, and activity by simplifying and imposing order. In that both traditions make implicit or explicit claims to truth (Boyd & Crawford, 2012; Hjørland, 2013) based on the order they impose (or, alternatively, "extract") and seek to aid decision-making it is not surprising that they share ethical consequences in their application in systems of access and organization.

We suggest that information studies, particularly through the scholarship of knowledge organization, can provide relevant theory and best practices to explain and address the current ethical challenges in AI and machine learning. Furthermore, we believe that by attending to the parallels between these approaches we might establish a broader, interdisciplinary field of scholars using different techniques to simplify and represent reality and knowledge, of which knowledge organization and machine learning are but two of the most salient examples. In fact, in almost all fields of science and technology and many social sciences researchers sometimes need to make simplifying assumptions that we suspect are relevant to the current discussion around automated methods and ethical considerations.

3. Bias

An area of urgent importance in AI for which knowledge organization can provide a long debate is the issue of bias. As in the example of how African-American voices are or are not represented in natural language models (Blodgett & O'Connor, 2017; Hovy & Sprout, 2016) and in sentencing recommendations in criminal cases (Angwin, Larson, Mattu, & Kirchner, 2016) scholars are attempting to track back unjust outcomes to identify (and rectify) the sources of bias. Within knowledge organization the inevitability of bias is a relatively new consensus position (Feinberg, 2007; Mai, 2010) arising from a rich history of critiquing and revising systems to ameliorate harmful effects. We identify parallels between the sources of bias scholars identify in knowledge organization and those relevant to machine learning approaches. For example, identification of the corpus—the collection of data points from which an AI system learns and bases its decisions—as the source of bias, as in the case of natural language processing that uses only one dialect of English, is similar to the role that literary warrant—or the decision to base classification decisions on the extant collection of literature—plays in the bias of systems such as the Library of Congress Subject Headings toward the historical (often white and male) canon of American literature (Olson, 2000). The way in which knowledge organization scholars have taken these ubiquitous examples of bias to move toward transparency and "responsible bias"

(Feinberg, 2007) is a possible path forward for machine learning, consistent with the recent critiques within the field (Garfinkel, Matthews, Shapiro, & Smith, 2017).

4. Conclusion

We examine the parallels and contrasts between knowledge organization and data science with a focus on the subfields of AI and machine learning in particular. We have identified some of these parallels above; contrasts include the dominant model of top-down or bottom-up analysis, the role of theory, and the role of domain experts or the crowd in determining truth. Each of these contrasts helps define the limits of shared concerns between these areas but also illuminates emerging priorities in knowledge organization. In this way, scholarship from AI and machine learning can shed light back on contemporary issues in knowledge organization, such as the increasing reliance on user-generated as opposed to expert knowledge (Yi & Mai Chan, 2009; Zhitomirsky-Geffet, Erez, & Judit, 2017).

Our analysis aims to both highlight and aid making connections between information studies and AI through the parallel practical and theoretical considerations and concerns within the fields of knowledge organization and machine learning. We believe that such an approach is not only productive for both fields in addressing emergent issues through reference to established discussion but also an alternative to seeing traditional knowledge organization and machine learning as competing approaches to contemporary challenges around decision making (e.g., providing organization and access to large collections). Both approaches belong to a broader category of 'simplifications' of reality and the intent of this analysis is to contribute to the development of theory across a host of fields, running the gamut from classification and natural language processing to simulation and data visualization. The current attention to AI is an opportunity for information studies scholars to participate in a cultural moment in which scholars and the public are engaged and concerned with what it means to reduce experience and reality to "information."

Reference List:

Agüera y Arcas, B., Todorov, A., & Mitchell, M. (2018, January 11). Do algorithms reveal sexual orientation or just expose our stereotypes? *Medium*. https://medium.com/@blaisea/do-algorithms-reveal-sexual-orientation-or-just-expose-our-stereotypes-d998fafdf477

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016, May 23). Machine bias. *ProPublica*. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Barocas, S., & Boyd, D. (2017). Engaging the ethics of data science in practice. *Communications of the ACM*, *60*(11), 23–25. http://doi.org/10.1145/3144172

Blodgett, S. L., & O'Connor, B. (2017). Racial disparity in natural language processing: A case

study of social media African-American English. (Pre-print). https://arxiv.org/abs/1707.00061

Boyd, D., & Crawford, K. (2012). Critical questions for big data. *Information, Communication & Society*, *15*(5), 662–679. http://doi.org/10.1080/1369118X.2012.678878

Crawford, K., & Calo, R. (2016). There is a blind spot in AI research. *Nature*, *538*(7625).

Feinberg, M. (2007). Hidden bias to responsible bias: an approach to information systems based on Haraway's situated knowledges. *Information Research*, *12*(4).

Garfinkel, S., Matthews, J., Shapiro, S. S., & Smith, J. M. (2017). Toward algorithmic transparency and accountability. *Communications of the ACM*, *60*(9), 5–5. http://doi.org/10.1145/3125780

Hovy, D., & Spruit, S. L. (2016, August). The social impact of natural language processing. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics* (Vol. 2, pp. 591-598).

Hjørland, B. (2013). Theories of knowledge organization—theories of knowledge. *Knowledge Organization*, *40*(3), 169–181.

Levin, S. (2017a, September 9). LGBT groups denounce "dangerous" AI that uses your face to guess sexuality. *The Guardian*. https://www.theguardian.com/world/2017/sep/08/ai-gay-gaydar-algorithm-facial-recognition-criticism-stanford

Levin, S. (2017b, September 12). Face-reading AI will be able to detect your politics and IQ, professor says. *The Guardian*. https://www.theguardian.com/technology/2017/sep/12/artificial-intelligence-face-recognition-michal-kosinski

Mai, J.-E. (2010). Classification in a social world: bias and trust. *Journal of Documentation*, *66*(5), 627–642. http://doi.org/10.1108/00220411011066763

Olson, H. A. (2000). Difference, culture and change: The untapped potential of LCSH. *Cataloging & Classification Quarterly*, *29*(1–2), 53–71.

Star, S. L., & Bowker, G. C. (2007). Enacting silence: Residual categories as a challenge for ethics, information systems, and communication. *Ethics and Information Technology*, *9*(4), 273–280. http://doi.org/10.1007/s10676-007-9141-7

Wang, Y., & Kosinski, M. (2017). Deep neural networks are more accurate than humans at detecting sexual orientation from facial images. (Preprint). http://doi.org/10.17605/OSF.IO/HV28A

Yi, K., & Mai Chan, L. (2009). Linking folksonomy to Library of Congress subject headings: an exploratory study. *Journal of Documentation*, *65*(6), 872–900. http://doi.org/10.1108/00220410910998906

Zhitomirsky-Geffet, M., Erez, E. S., & Judit, B.-I. (2017). Toward multiviewpoint ontology construction by collaboration of non-experts and crowdsourcing: The case of the effect of diet on health. *Journal of the Association for Information Science and Technology*, *68*(3), 681–694. http://doi.org/10.1002/asi.23686