Effects of data ambiguity and cognitive biases on the interpretability of machine learning models in humanitarian decision making

Paulus, David; de Vries, Gerdien; van de Walle, Bartel

Publication date
2019

Document Version
Final published version

Published in
AI for Social Good - AAAI Fall Symposium 2019

Effects of data ambiguity and cognitive biases on the interpretability of machine learning models in humanitarian decision making

David Paulus[1], **Gerdien de Vries**[2] and **Bartel Van de Walle**[3]
[1]HumTechLab, Delft University of Technology
email: d.paulus@tudelft.nl
[2]Organization and Governance Section, Delft University of Technology
email: G.deVries-2@tudelft.nl
[3]HumTechLab, Delft University of Technology
email: B.A.vandeWalle@tudelft.nl

Abstract

The effectiveness of machine learning algorithms depends on the quality and amount of data and the operationalization and interpretation by the human analyst. In humanitarian response, data is often lacking or overburdening, thus ambiguous, and the time-scarce, volatile, insecure environments of humanitarian activities are likely to inflict cognitive biases. This paper proposes to research the effects of data ambiguity and cognitive biases on the interpretability of machine learning algorithms in humanitarian decision making.

1 Introduction

Humanitarian response comprises a wide range of activities conducted by a multitude of actors in diverse contexts [1]. Activities include the search and rescue of wounded and deceased, delivery of aid to the affected, camp management and inter-organizational coordination [23]. Actors are the affected communities themselves, local and national groups and organizations, governmental agencies, the private sector, the military, international non-governmental organizations and the United Nations, as well as digital volunteers [2]. Humanitarian contexts can be categorized through locus, type and extent of disasters and crises as well as by the social, cultural and political environments they occur in [12].

Implementing organizations of humanitarian activities mainly operate through funds from donor country governments [5]. From that perspective, allocation decisions are being made on three levels: on a donor level regarding if and what amounts of funds are to be allocated to what humanitarian context and recipient organization. On an organizational headquarter level regarding what resources are to be allocated to what field operations. And on a field level regarding what resources are to be allocated to local partners and how to allocate aid to groups of affected people.

On all three levels, decision makers are challenged by data ambiguities and are influenced by cognitive biases but have to make decisions anyway. Often, decisions are made in the absence of reliable data and - for example due to stress, time-, and resource-constraints - under the influence of cognitive biases [3].

An example case

Yemen experiences the worst humanitarian crisis of today [25]. Activities for emergency food assistance have received the largest funds from international donor countries compared to activities in other sectors (e.g. health, education, protection) in Yemen (Figure 1).

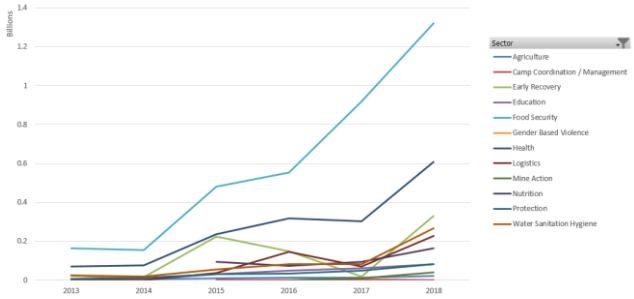

Figure 1. Funds donated to humanitarian sectors in Yemen. In USD. Source: UN OCHA Financial Tracking Service.

But funds have not been sufficient. And in cases where humanitarian organizations cannot reach all affected people, they have to prioritize certain groups and areas over others [28].

To support allocation decisions, machine learning can support the analysis of population data in combination with food security indicators and data on the changing conflict in Yemen. Ideally, algorithms predict how many rations will be needed when and where, thereby increasing the number of people reached, mitigating human suffering and reducing operational costs [17].

The efficiency of a machine learning algorithm depends on the data it is trained with [20]. If the data is not accurately representing reality or if it is lacking important attributes, these shortcomings are cascading into the model and its predictions [22]. In humanitarian response, data is often problematic [19]. But even if the data is reliable, biases can lead decision makers to ignore or misinterpret the re-

sults of the algorithm, or even collect and input data in a biased way. The time-scarce, stressful situation in Yemen is likely to inflict cognitive biases that can affect decision makers in e.g. deciding over whether to conduct a needs assessment, what data to trust and how to act on a prediction of a model.

2 Previous work

Data ambiguity has been defined as one aspect of data quality that leads to false interpretations because of inaccuracies within the data [6]. Data ambiguity is thus one of a number of aspects that make up data quality. [26] proposed a conceptual framework of data quality in which they categorized data characteristics into four categories: intrinsic, contextual, representational and accessibility data quality. Ambiguity, in general terms, is the quality of being open to more than one interpretation, and closely related to inexactness. This paper uses the term data ambiguity to summarize the data quality characteristics of accuracy, relevancy, timeliness, completeness and puts a particular focus on interpretability.

Interpretability of data for humanitarian organizations was studied by [4], in their study on decision makers' information needs in the response to Typhoon Haiyan. The authors argued that the heterogeneity of data caused confusion between different organizational levels and suggested more standardization of data and a comprehensive ontology that enables a "shared conceptualization" of the disaster situation. They further stressed the role of coordination to provide "relevant, accurate and timely" data to all actors in the response. They found "necessary but competing agendas" between organizations' headquarters and field offices. Headquarters required frequent, accurate reports which field staff could not deliver and which impeded their own planning and activities on the ground. Their findings show a strong effect of data ambiguity on intra-organizational understanding.

The lack of reliable data and the uncertainty it creates, leaves room for heuristics and cognitive biases humanitarian actors fall back to in their decision making processes. The idea of cognitive biases and that they influence individuals through heuristics so that their judgments differ from purely rational thinking, was introduced by [10]. Biases are often characterized as a byproduct of information processing limitations: because of a lack of time or capacities, people use these mental shortcuts to judge and decide. There is "broad consensus that human decision making relies on a repertoire of simple, fast, heuristic decision rules to be used in specific situation" [8, p. 729]. While most studies on cognitive biases label them as fallacies, [7] showed that simple decision making approaches can perform equally or better compared to sophisticated approaches that try to gather and process all available information. And according to [8], many biases may have developed to favor "inexpensive, frequent errors rather than occasional disastrous ones" [8, p. 731].

While cognitive biases in humanitarian emergencies have only been hypothesized yet, evidence exists for cognitive biases in other high-risk situations. Examples are studies on group behavior and decision making in physically and mentally extremely challenging events, for example [21, 15]. From those studies, some biases appear more dominant than others and include sunk cost fallacy, overconfidence, recency bias and confirmation bias. These are also part of the list of biases postulated by [3] that might be influencing decision makers in humanitarian emergencies. [3] argues these biases might be influential because of the "stress and pressure, distorted, lacking and uncertain information" in humanitarian emergencies. And that decision makers in the humanitarian sector develop coping strategies to deal with the high number of decisions they have to conduct in short periods of time.

To the best of our knowledge, effects of data ambiguity and cognitive biases on the interpretability of machine learning models in humanitarian response have not been studied. So far, machine learning has been applied and studied in a number of applications within the humanitarian sector. [19] found that high accuracy levels can be achieved by applying machine learning models on mapping tasks for refugee settlements. And [16] proposed a hybrid human-machine learning approach to analyze large amounts of satellite imagery data in disaster contexts. And a number of scholars utilized machine learning approaches to assess social media data during emergencies [9]. [22] discussed cognitive biases and their potential effects on the interpretations by human analysts of rule-based machine learning models. And [24] found that machine learning algorithms can perform better in classifying small datasets when biases are artificially coded into the algorithms.

We therefore propose a set of research questions to investigate the effects of data ambiguity and cognitive biases on the interpretability of machine learning models in humanitarian decision making in the next section.

3 Proposed Research

Looking back at the three-level perspective on the humanitarian sector, and taken into account the evidence of conflicting understandings between these levels, the proposed research should investigate the following questions for each level individually.

What is a suitable measure for interpretability [18] of machine learning models in humanitarian decision making?
Machine learning analysts and decision makers in donor country agencies likely follow different rules to achieve different results than their counterparts in recipient organizations' headquarters and field offices. Also their professional backgrounds and trainings might vary, leading to different approaches to operationalize machine learning models and assess and value their results.

What characteristics of data ambiguity influence interpretability of machine learning models in humanitarian decision making?
Data characteristics – including accuracy, timeliness, completeness, trustworthiness, format – might be valued differently throughout the three levels and are highly context-dependent. During search and rescure operations, decision

makers might favor fast results over completeness and accuracy, to save time and accelerate relief operations.

What cognitive biases influence interpretability of machine learning models in humanitarian decision making?
Above mentioned literature points to potential biases that analysts and decision makers might be prone to adopt, which can have positive and negative effects on interpretability. Studies within cognitive psychology, however, have found many more biases that are worthwhile to investigate in humanitarian decision making.

What interventions support positive effects and mitigate negative effects on the interpretability of machine learning models in humanitarian decision making?
Effects that are discovered through the previous questions, might then be available to be strengthened or weakened through individual or organizational interventions. Sensemaking [27] was previously suggested as an organizational measure to adapt to uncertain and ambiguous humanitarian environments [14]. Debiasing, which often entails a training component on personal awareness, also holds potential to interfere with either positive or negative effects on interpretability [13, 11].

References

[1] Balcik, B., Beamon, B. M., Krejci, C. C., Muramatsu, K. M., & Ramirez, M. (2010). Coordination in humanitarian relief chains: Practices, challenges and opportunities. International Journal of Production Economics, 126(1), 22-34.

[2] Burns, R. (2015). Rethinking big data in digital humanitarianism: Practices, epistemologies, and social relations. GeoJournal, 80(4), 477-490.

[3] Comes, T. (2016). Cognitive biases in humanitarian sensemaking and decision-making lessons from field research. In 2016 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support, CogSIMA 2016 (pp. 56–62). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/COGSIMA.2016.7497786

[4] Comes, T., Vybornova, O., & Van de Walle, B. (2015). Bringing Structure to the Disaster Data Typhoon: An Analysis of Decision-Makers' Information Needs in the Response to Haiyan. In 2015 AAAI Spring Symposium Series. Retrieved from https://www.aaai.org/ocs/index.php/SSS/SSS15/paper/view/10288

[5] Development Initiatives. (2018). Global Humanitarian Assistance Report 2018. Global Humanitarian Assistance.

[6] Eppler, M. J. (2006). Managing Information Quality: Increasing the Value of Information in Knowledge-intensive Products and Processes. Berlin, Heidelberg: Springer-Verlag.

[7] Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. Psychological Review. US: American Psychological Association. https://doi.org/10.1037/0033-295X.103.4.650

[8] Haselton, M. G., Nettle, D., & Andrews, P. W. (2005). The evolution of cognitive bias. In D. M. Buss (Ed.), The Handbook of Evolutionary Psychology (pp. 724–746). New York, US: John Wiley & Sons Inc.

[9] Imran, M., Castillo, C., Diaz, F., & Vieweg, S. (2015). Processing social media messages in mass emergency: A survey. ACM Computing Surveys (CSUR), 47(4), 67.

[10] Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. Cognitive Psychology, 3(3), 430–454. https://doi.org/10.1016/0010-0285(72)90016-3

[11] Koehler, D. J., & Harvey, N. (Eds.). (2004). Blackwell handbook of judgment and decision making. Blackwell handbook of judgment and decision making. Malden: Blackwell Publishing. https://doi.org/10.1002/9780470752937

[12] Leaning, J., & Guha-Sapir, D. (2013). Natural disasters, armed conflict, and public health. New England journal of medicine, 369(19), 1836-1842.

[13] Morewedge, C. K., Yoon, H., Scopelliti, I., Symborski, C. W., Korris, J. H., & Kassam, K. S. (2015). Debiasing Decisions: Improved Decision Making With a Single Training Intervention. Policy Insights from the Behavioral and Brain Sciences, 2(1), 129–140. https://doi.org/10.1177/2372732215600886

[14] Muhren, W. J. (2011). Foundations of sensemaking support systems for humanitarian crisis response. University of Tilburg.

[15] National Research Council. (2015). Measuring Human Capabilities: An Agenda for Basic Research on the Assessment of Individual and Group Performance Potential for Military Accession. Washington, DC: The National Academies Press. https://doi.org/10.17226/19017

[16] Ofli, F., Meier, P., Imran, M., Castillo, C., Tuia, D., Rey, N., ... & Joost, S. (2016). Combining human computing and machine learning to make sense of big (aerial) data for disaster response. Big data, 4(1), 47-59.

[17] Peters, K., Fleuren, H., den Hertog, D., Kavelj, M., Silva, S., Goncalves, R., Ergun, O., Soldner, M. (2016). The Nutritious Supply Chain: Optimizing Humanitarian Food Aid. (CentER Discussion Paper; Vol. 2016-044). Tilburg: CentER, Center for Economic Research.

[18] Poursabzi-Sangdeh, F., Goldstein, D. G., Hofman, J. M., Vaughan, J. W., & Wallach, H. (2018). Manipulating and measuring model interpretability. arXiv preprint arXiv:1802.07810.

[19] Quinn, J. A., Nyhan, M. M., Navarro, C., Coluccia, D., Bromley, L., & Luengo-Oroz, M. (2018). Humanitarian applications of machine learning with remote-sensing data: Review and case study in refugee settlement mapping. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 376(2128). https://doi.org/10.1098/rsta.2017.0363

[20] Rebala, G., Ravi, A., & Churiwala, S. (2019). Introduction to Machine Learning. Cham, Switzerland: Springer Nature Switzerland AG. https://doi.org/10.1016/B978-0-12-396502-8.00013-9

[21] Roberto, M. A. (2002). Lessons from Everest: The Interaction of Cognitive Bias, Psychological Safety, and System Complexity. California Management Review, 45(1), 136–158. https://doi.org/10.2307/41166157

[22] Sengupta, E., Garg, D., Choudhury, T., & Aggarwal, A. (2019). Techniques to Elimenate Human Bias in Machine Learning. 2018 International Conference on System Modeling & Advancement in Research Trends (SMART), 226–230. https://doi.org/10.1109/sysmart.2018.8746946

[23] Stumpenhorst, M., Stumpenhorst, R., & Razum, O. (2011). The UN OCHA cluster approach: gaps between theory and practice. Journal of Public Health, 19(6), 587-592.

[24] Taniguchi, H., Sato, H., & Shirakawa, T. (2018). A machine learning model with human cognitive biases capable of learning from small and biased datasets. Scientific Reports, 8(1), 23–25. https://doi.org/10.1038/s41598-018-25679-z

[25] United Nations. (2018). Global Humanitarian Overview 2019.

[26] Wang, R. Y., & Strong, D. M. (1996). Beyond Accuracy: What Data Quality Means to Data Consumers. Journal of Management Information Systems, 12(4), 5–33. https://doi.org/10.1080/07421222.1996.11518099

[27] Weick, K. E., Sutcliffe, K. M., & Obstfeld, D. (2005). Organizing and the Process of Sensemaking. Organization Science, 16(4), 409–421. https://doi.org/10.1287/orsc.1050.0133

[28] World Food Programme. (2018). Annual performance report for 2017. Rome.