Socially Augmented Crowdsourced Collection of Folk Theories

Jonas Oppenlaender
University of Oulu
Oulu, Finland
jonas.oppenlaender@oulu.fi

Abstract

Folk theories represent the users' beliefs of how an AI-driven system works. Prior research investigated folk theories as a lens on users' reasoning about the algorithmically-curated content of news feeds on social media platforms, primarily in interview studies and surveys. In this work-in-progress, we present an interactive interface for the structured capturing and crowdsourced collection of folk theories at scale.

Introduction

Machine Learning (ML) has become a pervasive influence in our lives. YouTube, Twitter, and Facebook are examples of social media websites that use Machine Learning algorithms to curate, select and present information. Since the operation of the algorithm is typically opaque, users often develop theories about the algorithm in order to plan or reflect on their behavior – their *folk theories* (Eslami et al. 2016; DeVito et al. 2018). These intuitive theories are mental models representing the users' belief of how an opaque system or algorithm works (Rozenblit and Keil 2002; Jones et al. 2011; Gelman and Legare 2011). An understanding of an algorithm's inner workings may contribute to interpretability and Explainable Artificial Intelligence (XAI), two research areas that have received much attention in recent years (Biran and Cotton 2017).

Studying users on social media platforms and their beliefs about an opaque algorithm is a difficult problem for three reasons. First, social media platforms are dynamic in the sense that users, creators, and the algorithm influence each other. This dynamic and evolving interrelationship between the different key actors on the platform makes the platform difficult to study. Second, folk theories may exist in a state of explanatory co-existence (Shtulman and Lombrozo 2016). A ground truth for folk theories does not exist, and opposing theories may co-exist between or even within users. Third, as platforms monopolize and protect their data, information about the users and their decision process is difficult to obtain.

Prior studies investigated folk theories primarily on a limited scale with costly qualitative methods, such as interviews. Eslami et al. (2016), for instance, interviewed 40 Facebook users, and DeVito et al. (2018) interviewed 28 social media users. We lack insights into the users' reasoning and decision process on a larger scale. Perhaps the largest investigation to date with 3375 study participants investigated the news feeds of Facebook and Twitter using Wiki-surveys (French and Hancock 2017). The Wiki-survey method elicits preferences in pair-wise comparisons and also allows respondents to enter their own options. However, the property of explanatory co-existence of folk theories suggests that eliciting theories via pairwise comparisons may not be the best choice.

In this research project, we aim to further our understanding of how users form theories of interpreting AI decisions. We focus our investigation on YouTube, one of the largest social media websites. We will explore how people on YouTube reason about the algorithmic curation of their news feed and about the creators' content creation strategies.

Proposed Method

Folk theories should best be captured in their natural context (Rozenblit and Keil 2002). The question, thus, is how can folk theories be elicited *in situ* on the social media platform, and how should the theories be processed and stored?

Eliciting Folk Theories in situ

To capture the folk theories in situ and at scale, we posit there could be two methods:

1. *Audit studies* (Heckman and Siegelman 1993; Sandvig et al. 2014) are a method to probe an algorithm by eliciting responses from the algorithm.
2. *Annotations*: Users could be asked label elements visible on the screen.

Annotations pose the challenge that not all elements of a folk theory may be visible on the screen at a time. The algorithm could, for example, suggest a video because a certain action was taken by the user in the past. Audit studies, on the other hand, are well-aligned with the microtask paradigm on many crowdsourcing platforms. They also align with an active learning ML approach in which a model of the algorithm is built piece-by-piece from probing different examples. On the other hand, due to workers contributing anonymously on microtask crowdsourcing platforms, audit studies may not be suitable since it is not possible to retrieve an item from the worker's personal news feed.

Abstracting Folk Theories

Capturing folk theories as unstructured natural language would lead to difficulties with interpreting the collected theories due to different ways of phrasing the theories. French and Hancock (2017) opted to represent folk theories as conceptual metaphors. Similarly, Hope et al. (2017) abstracted analogies into schemas, and Kittur et al. (2014) used schemas to encode sensemaking processes.

In our project, we aim to identify and learn the simplified structural representations of the collected folk theories from the data provided by the crowd. Translating the folk theories into schemas will account for the structural similarity in the theories.

Capturing Folk Theories with an Interface

In our project, we will iteratively develop an interactive annotation interface for the crowdsourced collection of folk theories at scale. The interface will be available as a Chrome browser extension and a stand-alone version.

A first prototype of the interface is depicted in Figure 1. The interface will allow crowd workers and YouTube users to annotate YouTube's user interface with their set of folk theories. Data capture will be enriched with data from other users, following the model for socially augmented information foraging by Kittur et al. (2014).

The interface consists of a sidebar overlaid on top of the YouTube website. Users can select any video on the left part of the screen, and enter their theory of why this video was recommended to them in the sidebar. The interface will highlight relevant parts of the theories as they are typed, using parts-of-speech (POS) tagging, named entity recognition (NER), and information learned from the collected theories from other users. The folk theory will automatically be classified, using a suitable classification model, based on the theories collected from other users. Users can optionally assign the theory to an existing theory schema, or create a new schema with a simple schematic notation, such as <like-video>, <search-video>, and <search-topic>. Further, the user is asked to rate their confidence in the folk theory. The user can review and manage the list of theories in the interface.

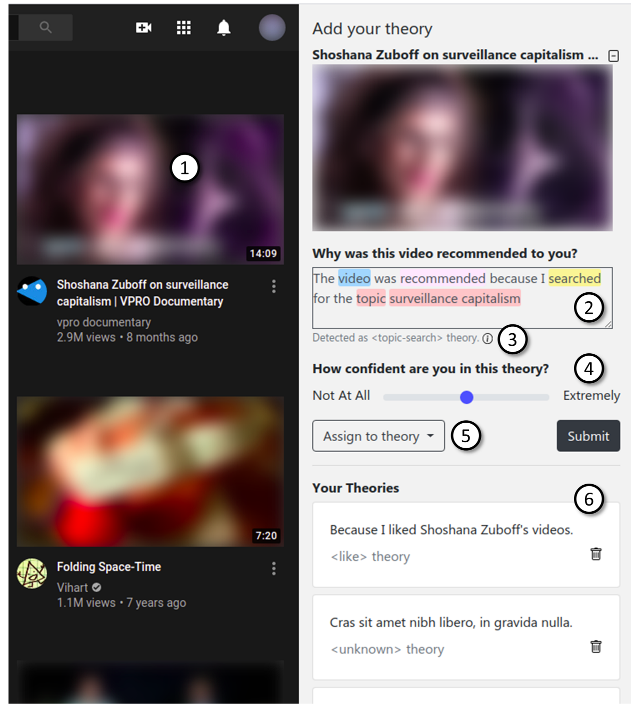

Figure 1: **Annotation interface** with means for entering folk theories ② for a selected video ①, detecting the theory ③, rating the confidence in the theory ④, assigning the video to existing theories ⑤, and reviewing a list of theories ⑥.

Future Work and Conclusion

The data-driven empirical investigation of users' folk theories has potential to shed light on how users reason about opaque algorithms on social media platforms. Our prototypical interface aims to capture the users' reasoning about the algorithm *in situ* on the social media platform. The data collection is socially augmented with data and schemas from other users.

As a next step, we will develop a solution for processing the collected theories with Natural Language Processing (NLP) methods. This includes the selection of a suitable ML model for classifying the folk theories and detecting the underlying schemas in folk theories.

In future work, the collected schemas could be used to dynamically learn and generate scaffolds from the user-provided folk theories to help users provide more useful and better structured folk theories. The scaffolds should guide the users in entering their folk theories, but without rigidly prescribing a schema. Our interface aims to create an awareness about different theories, but also allows users to enter new schematic representations of folk theories.

An understanding of how users reason about algorithmically-curated content could provide valuable insights into how web-based systems can be designed for fairness, accountability, and transparency. These three constituents contribute towards increased awareness and understanding of AI decisions, and could be a step stone to support the formation of trusting social relationships with Artificial Intelligence.

Acknowledgements

This work is supported by grants from the Finnish Foundation for Technology, the Tauno Tönning Foundation, the Jenny and Antti Wihuri Foundation, the Riitta and Jorma J. Takanen Foundation, and a scholarship from the Nokia Foundation.

References

Biran, O., and Cotton, C. V. 2017. Explanation and justification in machine learning: A survey. In *IJCAI 2017 Workshop on Explainable Artificial Intelligence (XAI)*.

DeVito, M. A.; Birnholtz, J.; Hancock, J. T.; French, M.; and Liu, S. 2018. How people form folk theories of social media feeds and what it means for how we study self-presentation. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, 1–12. New York, NY, USA: Association for Computing Machinery.

Eslami, M.; Karahalios, K.; Sandvig, C.; Vaccaro, K.; Rickman, A.; Hamilton, K.; and Kirlik, A. 2016. First i "like" it, then i hide it: Folk theories of social feeds. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, 2371–2382. New York, NY, USA: Association for Computing Machinery.

French, M., and Hancock, J. 2017. What's the folk theory? Reasoning about cyber-social systems. *SSRN*.

Gelman, S. A., and Legare, C. H. 2011. Concepts and folk theories. *Annual review of Anthropology* 40:379–398.

Heckman, J. J., and Siegelman, P. 1993. The urban institute audit studies: Their methods and findings. In Fix, M., and Struyk, R., eds., *Clear and Convincing Evidence: Measurement of Discrimination in America*. Urban Institute Press. 187–258.

Hope, T.; Chan, J.; Kittur, A.; and Shahaf, D. 2017. Accelerating innovation through analogy mining. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '17, 235–243. New York, NY, USA: Association for Computing Machinery.

Jones, N. A.; Ross, H.; Lynam, T.; Perez, P.; and Leitch, A. 2011. Mental models: An interdisciplinary synthesis of theory and methods. *Ecology and Society* 16(1).

Kittur, A.; Peters, A. M.; Diriye, A.; and Bove, M. 2014. Standing on the schemas of giants: Socially augmented information foraging. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*, CSCW '14, 999–1010. New York, NY, USA: Association for Computing Machinery.

Rozenblit, L., and Keil, F. 2002. The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive Science* 26(5):521–562.

Sandvig, C.; Hamilton, K.; Karahalios, K.; and Langbort, C. 2014. Auditing algorithms: Research methods for detecting discrimination on internet platforms. In *Data and Discrimination: Converting Critical Concerns into Productive Inquiry*.

Shtulman, A., and Lombrozo, T. 2016. Bundles of contradiction: A coexistence view of conceptual change. In *Core Knowledge and Conceptual Change*, Oxford series in cognitive development. New York, NY, USA: Oxford University Press. 53–71.