Causal Inference over Longitudinal Data to Support Expectation Exploration

Emre Kıcıman
Microsoft Research
emrek@microsoft.com

ABSTRACT

Many people use web search engines for expectation exploration: exploring what might happen if they take some action, or how they should expect some situation to evolve. While search engines have databases to provide structured answers to many questions, there is no database about the outcomes of actions or the evolution of situations. The information we need to answer such questions, however, is already being recorded. On social media, for example, hundreds of millions of people are publicly reporting about the actions they take and the situations they are in, and an increasing range of events and activities experienced in their lives over time. In this presentation, we show how causal inference methods can be applied to such individual-level, longitudinal records to generate answers for expectation exploration queries.

CCS CONCEPTS

• **Information systems** → *Web search engines*; *Specialized information retrieval*;

KEYWORDS

search, causal inference, social media, expectation exploration

ACM Reference Format:

Emre Kıcıman. 2018. Causal Inference over Longitudinal Data to Support Expectation Exploration. In *SIGIR '18: The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, July 8–12, 2018, Ann Arbor, MI, USA*. ACM, New York, NY, USA, 1 page. https://doi.org/10.1145/3209978.3210215

TALK DESCRIPTION

Everyone, at some point in their lives, finds themselves in an unfamiliar situation, considering what they should do, and trying to understand what to expect of the future. We see evidence of such *expectation exploration* in web searches, with people seeking the possible consequences of their choices and the outcomes of situations. These explorations cover both consequential topics, such as life-changing education and career choices (e.g., "Should I join the military?") or major financial and personal decisions (e.g., "Should I move to California?"); as well as more quotidian topics, such as the consequences of purchase decisions, athletic training regimens and dating rituals.

The answers to these questions are not readily available in a knowledge base or Wikipedia. But, the information necessary to answer these questions is already being recorded on social media, where hundreds of millions of individuals regularly and publicly report their personal experiences, including the situations they are in, the actions they take, and the experiences they have afterwards. For example, people talk about work or relations health and dietary practices, and even log information about their illnesses and coping strategies. People report and share this information for many reasons: keeping in touch with friends, gaining social capital, diary-keeping, or even helping others. And with increasing use of personal sensors and devices, from exercise trackers to health monitors, such data streams are becoming more regular, more detailed and more reliable. These longitudinal data streams, in aggregate, capture a rich set of relationships between the situations in which people find themselves, the actions they choose to take, and the outcomes they experience.

In this presentation, we describe how such individual-level, longitudinal datasets can be analyzed with causal inference methods to directly identify what can be expected following some action or individual experience. The result is a semi-structured *expectation map* that captures how situations are likely evolve over time. These expectation maps can be use as an important building block for a wide variety of data-driven search, decision-support and forecasting applications—from automatically generating decision aids, such as pros and cons lists, to helping individuals ground their own experiences in how a situation is likely to evolve over time. In addition, expectation maps may be useful for policy makers' and scientists' explorations across a variety of domains.

Acknowledgements

This presentation brings together ideas developed through close collaboration and discussion with many others. In particular, I would like to thank Paul Bennett, Scott Counts, Munmun De Choudhury, Susan Dumais, Adam Fourney, Alexandra Olteanu, Matt Richardson, Jorgen Thelin, Onur Varol, Ryen White, and Lidong Zhou.

SIGIR '18, July 8–12, 2018, Ann Arbor, MI, USA

ACM ISBN 978-1-4503-5657-2/18/07.
https://doi.org/10.1145/3209978.3210215