Towards Extracting Explanations from Mobile Data

Api Hasthanasombat
ah953@cam.ac.uk
University of Cambridge

Cecilia Mascolo
cm542@cam.ac.uk
University of Cambridge

MOTIVATION

Existing work with mobile data sets have largely focused on prediction and recognition. For instance, estimating venue closure based on previous neighbourhood mobility patterns from a mobile social network [1], predicting users' context and activity based on sensor data available on a mobile device [4], to predicting the evolution of disease from human mobility data [6].

Often however, we are interested in answering explanatory questions. Is the existence of a certain type of venue responsible for the decrease in footfall for another neighbouring venue? Do users who consistently report higher levels of alertness do so because they have higher daily activity levels? Or does the existence of certain types of venue influence the health outcome of an area? The main distinction between answering these explanatory questions, as opposed to prediction or identification, is that it takes us closer to devising interventions on the necessary variables to influence the target outcome - whether it be venue closure or disease evolution - closer to the desired result.

There are challenges that come with answering these explanatory questions. The main issue is that data from these mobile systems are *observational*, meaning that they did not come from any controlled experiment. This means that attempts to elicit causation must look further than correlation; there may be several variables that confound the influence of one another on the target outcome, which in turn could also influence each other. However, under a certain set of assumptions and in combination with other sources of observational data, we may be able to tackle this problem.

RELATED WORK

The main toolkit used to approach this problem comes from a growing literature in causal inference methodology. Roughly, these methods can be put into two categories: inferring the magnitude of an effect when we already know the structure of how variables influence each other, and inferring the structure itself - called causal discovery. The structure is usually represented by a Bayesian Network, often called a Causal Bayesian Network (CBN), or more recently, a Structural Causal Model (SCM) [3].

Attempts to use these tools for explanatory questions in the mobile literature is under-explored. Existing work has started to look at how certain treatments, such as the amount of exercise and sociability measured by sensors, affect self reported outcomes such as stress [5]. Or attempts to find the effect of certain types of venues on health outcomes, such as sporting facilities on antidepressant prescriptions, using crowd-sourced data [2].

However, many significant issues remain for valid inferences under these circumstances. First and foremost is the problem with confounding; in most mobile scenarios, we do not have complete knowledge of the causal structure at hand. We can make strong assumptions but this decreases the significance of our inferences. Additionally, evaluating the inference is not straightforward, as there is no 'ground truth' - this is what we are searching for in the first place. This means we have to result to using synthetic or semi-synthetic data: data that has been constructed based on the observed dataset, where we are in control of the 'true effect'.

OUTLINE

The presentation is about post-hoc analysis and its potential with the increase in observational data, a large portion of which comes from mobile devices.

We then talk about the challenges that come with this type of analysis, and the currently available tools to tackle them, along with the required assumptions.

Specifically, we look at the graphical representation of confounding, and the two main strategies that can be used to resolve it, and some specific techniques that fall into these two categories. We show that there is no way to estimate causal effects without the use of assumptions.

We will then look at the problem that comes with trying to evaluate these estimates.

REFERENCES

[1] D'Silva, K., Jayarajah, K., Noulas, A., Mascolo, C., and Misra, A. The role of urban mobility in retail business survival. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 2*, 3 (2018), 100.

[2] Hasthanasombat, A., and Mascolo, C. Understanding the Effects of the Neighbourhood Built Environment on Public Health with Open Data. 11.

[3] Pearl, J. *Causality*. Cambridge University Press, Sept. 2009. Google-Books-ID: f4nuexsNVZIC.

[4] Radu, V., Tong, C., Bhattacharya, S., Lane, N. D., Mascolo, C., Marina, M. K., and Kawsar, F. Multimodal Deep Learning for Activity and Context Recognition. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 1*, 4 (Jan. 2018), 157:1–157:27.

[5] Tsapeli, F., and Musolesi, M. Investigating causality in human behavior from smartphone sensor data: a quasi-experimental approach. *EPJ Data Science 4*, 1 (Dec. 2015), 24.

[6] Xie, X., Wang, Y., Zhou, X., Mascolo, C., Noulas, A., and Chen, E. Predicting the Spatio-Temporal Evolution of Chronic Diseases in Population with Human Mobility Data. 3578–3584.

MobiUK'19, 1–2 July, Oxford, UK