2023 IEEE International Conference on Big Data and Smart Computing (BigComp) | 978-1-6654-7578-5/23/$31.00 ©2023 IEEE | DOI: 10.1109/BigComp57234.2023.00088

Causal Analytic Process for Mobile Health Data

Gyuwon Jung, Sangjun Park, Uichin Lee*
School of Computing
KAIST
Daejeon, South Korea
{*gwjung, sangjun, uclee*}@*kaist.ac.kr*

Eun-Yeol Ma, Heeyoung Kim*
Department of Industrial and Systems Engineering
KAIST
Daejeon, South Korea
{*eyma1127, heeyoungkim*}@*kaist.ac.kr*

***Abstract*—This paper proposes a process of causal inference using data from mobile devices to better understand the relationships between human behaviors and contexts. A simple case study is presented based on the proposed method and shows the existence of causality in a given sample scenario.**

***Keywords*-Causal Inference; Mobile Data; Human Behavior;**

I. Introduction

The data collected daily via mobile devices such as smartphones and wearables (hereafter, "mobile data") have a huge potential for understanding human behaviors and getting new insights into them. Recent studies [1], [2] have investigated the causal relationships between human activities, mental states, and contexts by leveraging mobile data but lacked explanations of how to conduct causal analysis. This study proposes an analytic process of causal inference with mobile data, which can be utilized when randomized trials are not available (e.g., observational studies). Also, we demonstrate a simple case study based on the proposed method.

II. Causal Analytic Process

The causal inference process begins with setting up a causal scenario (i.e., hypothesis) by determining treatment and outcome variables. Next, mobile data are preprocessed to extract behavior and context features representing the variables. The features are divided into time windows to generate samples consisting of events within a given interval.

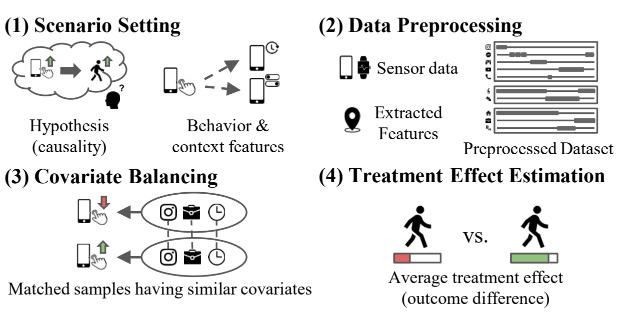

Figure 1. The overview of a causal analytic process for mobile data

Note that features other than the treatment and outcome are also included in the analysis if they are related to the causal scenario.

Then, the covariates that may affect the causal analysis are balanced and the hypothetical control and treated groups are generated based on the treatment level. Prior studies suggested matching methods [3] that match samples with similar covariates but different treatment levels to minimize the effects of covariates. After the matching process, the causal effect on the difference in the outcomes between the two hypothetical groups is estimated to determine the existence and strength of the causal relationship.

III. Case Study

We conducted a causal inference process on a single person's seven-day mobile dataset to test whether his/her social app usage was causally related to the sedentary duration. We extracted features including the location, time, sedentary state, and app usage. We also employed CEM [4] to balance the covariates using linear regression to estimate the causal effect.

In this process, the covariate distribution was well-balanced between the two hypothetical groups, with standardized mean differences of all covariates less than 0.25. In addition, a t-test on the coefficient of the treatment variable after linear regression showed a statistically significant difference in the outcomes (t-statistic $= .051$, p-value $< .05$), implying that this person's social app usage *causes* longer sedentary behavior.

IV. Conclusion

In this study, we proposed a causal inference process for mobile data in analyzing human behaviors. This process can be adopted in validating the effect of a mobile app (e.g., digital therapeutics) on the target behavior particularly when an observational study is the only option.

Acknowledgment

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Korea government (MSIT) (2020R1A4A1018774, 2022M3J6A1063021).

978-1-6654-7578-5/23/$31.00 ©2023 IEEE
DOI 10.1109/BigComp57234.2023.00088

REFERENCES

[1] F. Tsapeli and M. Musolesi, "Investigating causality in human behavior from smartphone sensor data: a quasi-experimental approach," *EPJ Data Science*, vol. 4, no. 1, p. 24, 2015.

[2] A. Mehrotra, F. Tsapeli, R. Hendley, and M. Musolesi, "My-traces: Investigating correlation and causation between users' emotional states and mobile phone interaction," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 3, pp. 1–21, 2017.

[3] E. A. Stuart, "Matching methods for causal inference: A review and a look forward," *Statistical science: a review journal of the Institute of Mathematical Statistics*, vol. 25, no. 1, p. 1, 2010.

[4] S. M. Iacus, G. King, and G. Porro, "Causal inference without balance checking: Coarsened exact matching," *Political analysis*, vol. 20, no. 1, pp. 1–24, 2012.