Capturing medical practice with indication embeddings

Rachel D. Melamed, PhD

1Biomedical Data Science, University of Chicago, Chicago, IL, USA

Background

The electronic health record is a rising resource for quantifying medical practice and discovering adverse effects of drugs. One of the challenges of health care data is the high dimensionality of the health record. Any study of patterns in health data must account for tens of thousands of potentially relevant diagnoses or treatments. In particular, comparative cohort studies are a major tool for assessing drug safety, but these studies require appropriate adjustment for confounding variables. Propensity score methods adjust for confounders by estimating the effect of these confounders on treatment assignment. One of the challenges of applying these methods to health care data is the high dimensionality of the health record, with tens of thousands of potential confounders. Thus, recent studies have proposed improvements to the propensity score, including large-scale regularized estimation, and automatic determination of relevant confounders (1–3). In this work, we develop indication embeddings, a way to reduce the dimensionality of health data while capturing the information relevant to treatment decisions.

Approach

We describe a new strategy to create embeddings that represent the relationship between medical history and the first prescription of a drug. This could include diseases that are indications for prescription of a drug, or drugs that act to ameliorate the side effects of other drugs (such as potassium supplements for diuretic prescriptions), or procedures that require prescription of a drug (such as colonoscopy preparations). In particular, we assess the performance of our embeddings for identification of de facto drug indications. While databases such as UMLS (4) and MEDI (5) codify the most well-established uses for drugs, these may not reflect off-label uses and changes in medical practice. Our method provides a data-driven way to identify such prescription choices.

Results

We demonstrate that these embeddings recover therapeutic uses of drugs. Then we use these embeddings as an informative representation of relationships between drugs, between health history events and drug prescriptions, and between patients at a particular time in their health history. We show the application of these embeddings for drug safety studies. In retrospective cohort studies, our low-dimensional representation helps in finding appropriate comparator drugs and constructing comparator cohorts.

Conclusions

Our embeddings create a representation of medical history centered on provider treatment choices. We expect that our approach will enable researchers to more rapidly conduct cohort studies and discover characteristics of medical practice.

References


