On the Promise of Topic Models for Abstracting Complex Medical Data: A Study of Patients and their Medications

Jenna Wiens
MIT
Cambridge, MA 02139
jwiens@mit.edu

John V. Guttag
MIT
Cambridge, MA 02139
guttag@csail.mit.edu

Eric Horvitz
Microsoft Research
Redmond, WA 98052
horvitz@microsoft.com

Abstract

Identifying similarities among patients is an integral part of both clinical practice and medical research. In principle, the proliferation of detailed electronic medical records should facilitate the growth of personalized medicine by making it possible to find useful cohorts of patients. However, the high dimensionality of modern medical records makes finding useful cohorts a challenge. Expert knowledge is often required to construct useful abstractions. We explore the use of topic models to discover such abstractions automatically. We test the proposed methods on the task of inferring useful abstractions from a list of thousands of medications. We investigate two different ways of generating relevant topic models, one leveraging Web content and another directly from electronic patient records. Applied to a corpus of 25,000 patient visits, we evaluate the potential utility of these topic models for computing relative patient similarity, and for predicting adverse outcomes at the next hospital admission, such as death or an abnormally long stay. We found that the classifiers built using the learned abstractions outperformed classifiers learned using an expert-devised drug classification scheme that is employed currently as an industry standard.