Learning Models for Monitoring and Prognosis from Electronic Health Data

Abstract:
Healthcare spending is nearing $3 trillion per year, but in spite of this expenditure, the US is outpaced by most developed countries with regard to outcomes. Until recently, one of the key bottlenecks for research in care delivery was the lack of data to analyze the health system’s workings. But today, post the HITECH in 2009, much of an individual’s health data is stored electronically. This opens up a wealth of opportunities for computational scientists.

In this talk, I will develop and solve two problems. The first is the challenge of accurate prognoses in chronic diseases where individuals show a great deal of variability. Here, access to accurate prediction tools can help tailor treatment decisions. We develop a probabilistic model that exploits the concept of subtypes to individualize predictions of disease trajectories. These subtypes identify groups of individuals that share a similar disease course. These subtypes are learned automatically from data. On a new individual, our model incorporates static and time-varying markers to dynamically update predictions of subtype membership and provide individualized predictions of disease trajectory.

The second challenge pertains to learning severity scores in the acute setting for scalable automated monitoring. For example, can we detect individuals declining due to impending adverse events early? The most commonly used approaches train a predictive model using retrospective data. But, this approach produces incorrect results due to interventional confounds (Paxton et al., 2013). We propose a novel ranking based solution for measuring the severity of an individual's latent health state. The learned score outperforms several state of the art clinical scores.

Time permitting; I will give an overview of other recent work!

We've tailored Netflix recommendations and Amazon shopping carts but we're only at the early stages of knowing how to use data to tailor health decisions.