CPSC 490: Senior Project Proposal
Annie Cook (alc98) MC ’17

Advisors

Advisor: Professor Ruzica Piskac, Computer Science
Additional: Abhishek Chandra, Yale College ‘16
Adam Chekroud, Psychology PhD Candidate

Background

Machine learning is becoming an increasingly prevalent tool in medical decision making. Propelled by developments in processing power, memory, storage, and access to large amounts of data, computers are being asked to tackle increasingly complex learning tasks within medicine. Previously, doctors were tasked with remembering patient information and combining that with previous knowledge and experience to make decisions about diagnosis and treatment. These decisions were inherently subjective due to the shortcomings of human memory and the limited data sets from which these decisions were formed -- an average doctor will likely only see tens or at most hundreds of patients with shared symptoms per year. Using machine learning, we can train models on patient data from thousands of patients to make objective decisions.

Following first-episode psychosis, up to 50% of patients, develop unfavourable disease outcomes. This is despite established treatments for schizophrenia and awareness about multiple outcome predictors. These prognostic factors, which have been identified in group-level studies, include sociodemographic factors, disease course variables, treatment adherence and response, psychiatric comorbidity and functional
and cognitive deficits. Clinicians lack the tools to dictate how and which of these factors should be combined for prediction.

**The Algorithm**

Machine learning techniques have been recently applied to psychosis treatment selection, where there is a wealth of potential baseline predictors and outcomes of interest, and has been published in the Lancet Psychiatry (Koutsouleris et al., 2016). To develop the algorithm, data from The European First Episode Schizophrenia Trial was used. The data came from minimally treated patients with first-episode psychosis recruited across 50 European sites and treated over 52 weeks in an open-label randomized clinical trial of five broadly used antipsychotics. Machine learning was applied to sociodemographic, clinical and neurocognitive variables from the dataset.

Results of the study demonstrated feasibility of using pre-treatment clinical information to determine generalisable, individual prediction of treatment outcomes. Using the machine learning algorithm resulted in a 40% increase in prognostic certainty compared with pre-test outcome probabilities in our population. Development of the algorithm is proof of concept that machine learning can aid in this specific treatment selection.

**My Project**

The algorithm is only the first step for effective use of machine learning in a clinical setting. Large enough medical data sets and adequate learning algorithms have been available for many decades, and although there have been many papers applying machine learning algorithms to medical data, very few have contributed meaningfully to clinical care. A key factor in solving this discrepancy is successful implementation --
ease of use, integration with the clinical workflow and resulting meaningful information received for both the clinician and the patient.

My task is to design, develop and implement a web based user interface for the algorithm in the form of a questionnaire for clinicians to use. This will excellently employ knowledge and skills from my combined major of computer science and psychology as the first phase will be psychology focused -- the design and development of a user interface to effectively collect patient data from clinicians for the algorithm, followed by the computer science phase of implementation on the web and interaction with the algorithm.

The main challenges I will face will be involved in the implementation of the questionnaire on the web. For this task, user experience is key as the benefit of the algorithm relies on clinicians actually using it. Two clinics have requested to use this tool in randomized controlled trials later in 2017 so it has the potential to make a large difference soon. I will be able to meet with them during the development and implementation stages to learn a bit about their workflow in order to educate my coding up of the questionnaire. Additionally, another challenge will be dealing with security as the project must be HIPAA compliant to keep medical data secure.

Deliverables

- Identify the minimum necessary set of questions required for the algorithm
- Wireframe UI for the questionnaire
- Code static questionnaire templates
- Implement conditional logic for questions
- Add necessary security to make project HIPAA compliant
- Deploy and meet with clinicians for modifications