Using Machine Learning to Recommend Oncology Clinical Trials

Anasuya Das[1], Leifur Thorbergosson[1], Aleksandr Griogorenko[1], David Sontag[2], and Iker Huerga[1]
[1]Memorial Sloan Kettering Cancer Center, [2]MIT

Introduction. Clinical trials serve an important role in oncology, not only advancing medical science but also offering patients promising therapy before it is widely available. Memorial Sloan Kettering Cancer Center (MSK) conducts over 500 therapeutic trials at one time; most are focused on a single type of cancer (e.g. breast, lung) reflecting subspecialized nature of care. However, clinical trial accrual is a challenge as patient-trial matching is a slow and manual process. We address this challenge via a machine learning-powered clinical trial recommendation engine designed to be deployed at the point of care.

Contribution. Previous approaches have looked at the match between the text in clinical trial eligibility criteria and patient data to create patient-trial recommendations using hand-engineered rules (Zhang and Demner-Fushman, 2017; Ni et al., 2015). We propose a more generalizable framework where a patient-patient similarity metric is learned using commonly available Electronic Medical Record (EMR) data. Trial eligibility is inferred based on how similar a query patient is to those already enrolled on a clinical trial. Development of generalizable approaches to learning patient similarities is an important machine learning challenge in healthcare, as problems where the number of outcomes is large and highly imbalanced arise frequently, as is the case in patient-trial recommendation.

Cohort. Using MSKs Electronic Data Warehouse we extract data from 19,114 patients who were enrolled on 1,518 therapeutic clinical trials at MSK between 2004 and 2016. To learn the similarity metric, we create positive and negative pairs and train a binary classifier to distinguish the two. Positive pairs are constructed from patients enrolled on the same trial (e.g. trial A); negative pairs are composed of a patient enrolled on trial A and a patient who enrolled on another trial at the time trial A was available. To address significant variation in clinical trial size we oversample small trials and undersample large ones when constructing training batches. We derive an initial patient feature vector from demographic, billing, procedure, lab, medication, and genetic mutation data. We then construct a new feature vector by applying an element-wise AND operation on the binarized patient vector, resulting in only entries common between members of a pair to be non-zero; and an OR operation, resulting in dissimilar elements being non-zero. The resulting vectors are concatenated (doubling the feature dimension to 15,536).

Methods. The AND-OR feature vector is fed into an L1-regularized linear SVM. The distance to the hyperplane of the SVM is used as a similarity score. After L1 regularization, dimension of the AND-OR feature vector is reduced to 1,353 non-zero features. As a baseline model, we implement a rules-based system that extracts demographics, cancer type, stage, genetic mutations, and labs from text in eligibility criteria and matches them to corresponding structured and text-extracted patient data. Due to highly customized nature of this approach, we only implement it for breast cancer (38 trials and 626 patients).

Evaluation. Data is stratified by cancer type and trial size, randomly shuffled and split into train, validate and test sets at a 70/15/15 proportion. Train and validate sets are used for hyperparameter tuning (via grid search for optimal regularization constant); test set is used to evaluate the final model using two metrics. Patient Quantile Index (PQI): all relevant clinical trials are ranked for each patient using the learned metric. The PQI is the percentile rank at which the patients trial appears in the ranked list. The percentile ranks are averaged across all patients for a final score. Trial AUC (TAUC): for each trial, patients enrolled on the trial are considered as positives, other patients - as negatives. These patients are jointly ranked, and a trial AUC is calculated and averaged for a final score. To reflect the real-world setting, comparisons are only made to trials that were available at the time of enrollment and within a cancer type (e.g. breast cancer patients are not evaluated against lung cancer).

Results. Our algorithm outperforms the baseline on the breast cancer subset (PQI: 0.86 vs. 0.70; TAUC: 0.83: 0.70) and maintains its performance on a test set that includes all cancer types (PQI: 0.89; TAUC: 0.84).

Discussion. We plan to deploy the model as part of MSKs EMR and evaluate it in a clinical setting, presenting clinicians with a ranked list of clinical trials for a given patient and a ranked list of potentially eligible patients for a given clinical trial to facilitate decision-making.

Y Ni, J Wright, J Perentesis, J Lingren, L Deleger, L Kaiser, I Kohane, and I Solti. In- creasing the efficiency of trial-patient matching: automated clinical trial eligibility pre- screening for pediatric oncology patients. BMC Medical Informatics and Decision Making, 15(28), 2015. doi: http://doi.org/10.1186/s12911-015-0149-3.

Kevin Zhang and Dina Demner-Fushman. Automated classification of eligibility criteria in clinical trials to facilitate patient-trial matching for specific patient populations. Journal of American Medical Informatics Association, Feb 2017. doi: 10.1093/jamia/ocw176. PubMed PMID:28339690.