Implementing Real-Time Patient Level Predictions Using PLP Models

Scott Brown, Jon Duke, James Fairbanks, Christine Herlihy, Kausar Mukadam, Jason Poovey, Melissa Rost
Georgia Tech Research Institute, Atlanta GA

ABSTRACT: *We built a general framework over the OHDSI patient-level prediction package to publish an API for a predictive analytics model for C. Diff, a deadly bacterial infection. This model makes risk predictions for individual patients based within a precision medicine decision support tool with the ability to toggle off and on the inclusion of potential drugs or treatments. This tool can aid medical practitioners when prescribing drugs with potential complications. With this ability, consideration needs to be given to the quality of the model itself as that effects the quality of the predictions and the inability of the model to predict for new treatments not available in the training data.*

Background

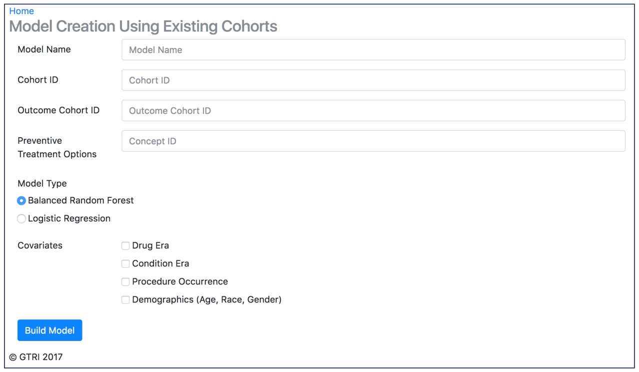
We used the OHDSI Patient Level Prediction (PLP) package to **extract patient-level data** from the MIMIC dataset comprising approximately 40,000 critical care patients. We were specifically interested in **building a predictive model** for Clostridium difficile, commonly known as C. Diff. This life-threatening infection in the colon is caused by taking antibiotics that kill the good gut bacteria so that bad bacteria can grow unchecked. In addition to the problems for the patient, costs attributable to C. Diff in the US are estimated at $6.3 billion annually and nearly 2.4 million days of inpatient stay annually so understanding the causes and factors involved more thoroughly has the potential to save lives and reduce a costly, resource-intensive problem.

The general PLP pipeline involves **creating a cohort of patients** in the database, **extracting features of interest** present in individual patients, and analyzing the resulting dataset by **building a predictive model**. The OHDSI PLP package to perform these tasks is written in R; however, the third step is implemented by having R call specific models in scikit-learn, a state of the art machine learning library written in Python. In order to facilitate access to the entire scikit-learn library, we use scikit-learn directly for the third step of analyzing the resulting data.

Integration with the Patient-Level Prediction Package

Index and outcome cohorts are first created using either dynamically constructed and executed SQL queries or OHDSI ATLAS. Next, we use the features defined in the PLP package to **extract patient-level information**. In order to support both **bulk queries for training** and **individual patient queries for predicting**, the features were extracted using QueryGarden SQL queries generated for both purposes. The resulting data from the cohorts, population, outcomes, and covariates is used in the analytics step.

Contact: melissa.rost@gtri.gatech.edu



Home

Model Creation Using Existing Cohorts

Model Name: Model Name

Cohort ID: Cohort ID

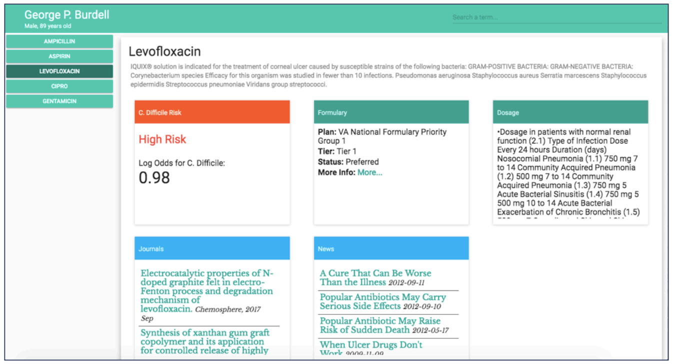
Outcome Cohort ID: Outcome Cohort ID

Preventive Treatment Options: Concept ID

Model Type
● Balanced Random Forest
○ Logistic Regression

Covariates
□ Drug Era
□ Condition Era
□ Procedure Occurrence
□ Demographics (Age, Race, Gender)

© GTRI 2017



George P. Burdell
Male, 89 years old

Levofloxacin

IQUIX® solution is indicated for the treatment of corneal ulcer caused by susceptible strains of the following bacteria: GRAM-POSITIVE BACTERIA: GRAM-NEGATIVE BACTERIA: Corynebacterium species Efficacy for this organism was studied in fewer than 10 infections. Pseudomonas aeruginosa Staphylococcus aureus Serratia marcescens Staphylococcus epidermidis Streptococcus pneumoniae Viridans group streptococci.

C. Difficile Risk

High Risk

Log Odds for C. Difficile:
0.98

Formulary

Plan: VA National Formulary Priority Group 1
Tier: Tier 1
Status: Preferred
More Info: More...

Dosage

•Dosage in patients with normal renal function (2.1) Type of Infection Dose Every 24 hours Duration (days) Nosocomial Pneumonia (1.1) 750 mg 7 to 14 Community Acquired Pneumonia (1.2) 500 mg 7 to 14 Community Acquired Pneumonia (1.3) 750 mg 5 Acute Bacterial Sinusitis (1.4) 750 mg 5 500 mg 10 to 14 Acute Bacterial Exacerbation of Chronic Bronchitis (1.5)

Journals

Electrocatalytic properties of N-doped graphite felt in electro-Fenton process and degradation mechanism of levofloxacin. *Chemosphere, 2017 Sep*

Synthesis of xanthan gum graft copolymer and its application for controlled release of highly

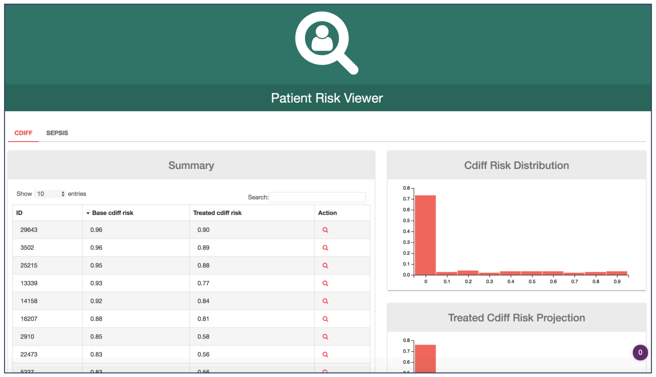
News

A Cure That Can Be Worse Than the Illness *2012-09-11*

Popular Antibiotics May Carry Serious Side Effects *2012-09-10*

Popular Antibiotic May Raise Risk of Sudden Death *2012-05-17*

When Ulcer Drugs Don't Work *2009-11-09*



Patient Risk Viewer

Summary

Show 10 entries   Search:

ID	Base cdiff risk	Treated cdiff risk	Action
29643	0.96	0.90	
3502	0.96	0.89	
25215	0.95	0.88	
13339	0.93	0.77	
14158	0.92	0.84	
16207	0.88	0.81	
2910	0.85	0.58	
22473	0.83	0.56	
5227	0.83	0.56	

Cdiff Risk Distribution

Treated Cdiff Risk Projection

Predicting Outcome Risk for a Single Patient Using a Trained Model

In order to predict C. Diff risk, we **build a predictive model** using scikit-learn. A web **user interface** (UI) allows researchers to choose any combination of features (i.e., drugs, conditions, procedures, and demographics) and model type (i.e., random forest or logistic regression) for training of this model and prediction.

A web application built on QueryGarden supports feature extraction over a REST API for this model. This extends the PLP pipeline of training a model and evaluating its accuracy by **enabling a real-time system to use the trained model to predict the risk of C. Diff for specific new patients.**

This capability is deployed in both the Advanced Clinical Decision Support (ACDS) system to provide precision medicine and the Patient Risk Viewer to summarize hospital-wide risk. For potential drugs to prescribe, the system **predicts an individualized expected patient outcome** to aid a physician in clinical decision-making.

Conclusions

Utilizing the OHDSI PLP package, we were able to extend the predictive model built from a cohort of patients in order to **provide a real-time prediction for new patients with the same selection criteria for the cohort**. Such a tool allows health practitioners the ability to gain knowledge of how an individual patient could be expected to react to a new drug or treatment before having to go through prescribing the drug and waiting for it to run its course.

This kind of advancement has the potential to revolutionize how medicine is prescribed; however, the medical community needs to be aware of potential flaws. The predictions are **only as good as the model** used to create them and creating the model is the real 'science' of the data science. Additionally, a predictive model only works with data that it was trained on, so **predicting for new covariates that weren't previously seen yields invalid results**. Precision medicine going forward will require collaboration between data scientists and medical professionals to understand the potential benefits and flaws.