<1>Chapter 10: Practicing techniques in context 2: Healthcare system redesign

Healthcare systems, particularly at the level of individual clinics, are undergoing dramatic workforce and financing changes, both in the United States and abroad. Team-based care, including non-physicians in care pathways, has become mainstream. Financing reforms to better fund preventive care, pro-active chronic disease management, and integrated care across physical and mental health domains, has also begun—in a dramatic shift from episodic outpatient care standards of the past.

In this Chapter, we tackle some of the key administrative and clinical questions that can arise during the course of healthcare system redesigns. We particularly practice the techniques of microsimulation and machine learning in the context of two current controversies in primary healthcare delivery reform.

<2>A microsimulation of behavioral health integration

A majority of patients with mild or moderate psychiatric diseases—stable depression, anxiety, and substance abuse, for example—are seen by a primary care physician, and are averse to seeing a psychiatrist to help them manage these diseases. As a result, many healthcare systems are planning for "behavioral health integration," or the integration of psychologists, social workers, counselors, or psychiatrists into the primary care setting. The idea behind behavioral health integration is to screen patients for psychiatric diseases when they are already visiting their primary care doctor for other conditions (e.g., diabetes, hypertension), and then be able to quickly create a treatment plan and referral to a psychiatric provider within the clinic, avoiding the "loss to follow-up" that often plagues psychiatric care services.

There are two common approaches for behavioral health integration, and our task is to understand how much an insurance company needs to pay a primary care clinic to ensure that the services are financially-viable for the clinic (that is, to ensure the clinic is paid sufficiently to not lose money when providing behavioral healthcare services). One approach, called the collaborative care model (CCM), is mostly telephone-based. It involves the primary care physician screening patients for conditions like depression, starting medication treatment for these conditions, then referring them to a nurse care manager in charge of keeping track of the referred patients and calling them to provide counseling and ensure they are reacting well to medication treatment. A psychiatrist has to review the cases periodically with the care coordinator. A second approach, called the primary care behaviorist (PCB) approach, involves the primary care physician arranging a same-day visit with an in-clinic psychologist or social worker, who provides coordinated clinic visits on the same days as primary care visits. A dedicated clinic space is needed for the visits. The differences between the two approaches are time and cost. The CCM model provides care mostly over the telephone, requiring time and salary from a nurse care manager and (periodically) a psychiatrist, but no in-clinic examination room space costs. The PCB approach provides care in person, requiring time and salary for a psychologist or social worker, as well as in-clinic examination room space costs.

Suppose our question is: what is the level of payment per visit that an insurer should provide (or that a clinic should accept) for a primary care practice to avoid losing money when undertaking behavioral health integration? Table 10.1 provides the necessary input variables for constructing a microsimulation model to answer this question (note, these are simulated data for teaching purposes, not real data).

[TABLE 10.1 HERE]

Let's construct a microsimulation model to estimate the distribution of costs a clinic might incur when implementing the CCM model, and when implementing the PCB approach. Let's then calculate how much the clinic needs to be paid per patient per year by the insurer to ensure the 95% confidence intervals of the net revenue for the clinic are positive.

First, we need to choose a suitably large number of simulations to conduct to ensure we gain a stable sense of the distribution of costs when repeatedly sampling from the distributions of all input parameter values. Let's say we want to simulate our clinic's cost 10,000 times. Next, we can use the `rnorm` function to create vectors for the number of patients who we expected to get referred to behavioral healthcare, and for each of the various costs under both the CCM and the PCB approaches:

```
      n = 10000
      patients=rnorm(n, mean  = 2000, sd = (2000-1500)/1.96)
      nursecaremanager = rnorm(n, mean = 100000, sd = (100000-
75000)/1.96)
      psychiatrist = rnorm(n, mean = 10000, sd = (10000-7500)/1.96)
      inclinicpsychsw = rnorm(n, mean = 120000, sd = (120000-
100000)/1.96)
      spacecosts = rnorm(n, mean = 1200, sd = (12000-1000)/1.96)
```

Essentially we have created 10,000 simulated realities of our clinic, sampling from the normal/Gaussian distributions for all our parameter values to get a sense of how much uncertainty we might have around each parameter estimate. Note that to get the standard deviations ('sd') for each distribution, we take the difference between the mean and lower 95% confidence interval and divide by 1.96 because the distance between the mean and 95% lower confidence interval is 1.96 times the standard deviation for a standard normal/Gaussian distribution.

Now we can compute the costs of the CCM approach and the PCB approach by adding up the vectors of costs for each. We can then estimate the distribution of payments needed per patient per year by dividing those costs by the number of patients per year:

```
ccmcosts = nursecaremanager + psychiatrist
pcbcosts = inclinicpsychsw + spacecosts

ccmcostperptperyr = ccmcosts/patients
pcbcostperptperyr = pcbcosts/patients
```

We can now examine the distribution of these costs and estimate how much the clinic needs to be paid per patient per year by the insurer to ensure the 95% confidence intervals of the net revenue for the clinic are positive:

```
> quantile(ccmcostperptperyr,c(.025,.975))
    2.5%    97.5%
39.17059 78.55240
> quantile(pcbcostperptperyr,c(.025,.975))
    2.5%    97.5%
44.54821 84.95072
```

We see that the minimum payment per patient per year for the CCM model is around $39, while the minimum payment for the PCB model is $45 per patient per year. The distributions of costs from our microsimulation are displayed in Figure 10.1.

[FIGURE 10.1 HERE]

<2>A machine learning model to recommend early intensive patient support

Receiving behavioral healthcare services through a telephone visit with a nurse, or even an in-person visit with a psychologist or social worker, can be extremely helpful for many patients. But for a handful of patients, these services are not sufficient. Some patients have such serious psychiatric disease that they need more intensive support through direct visits with a psychiatrist or even a brief inpatient hospitalization at a psychiatric hospital. Suppose that the behavioral health integration program has been instituted over the past year, and that we have data on the first 2,000 patients who were referred through the system. Among those patients, suppose a majority did very well in the program, but a subset did not, and required later referral to a psychiatrist. Could we come up with a machine learner that could predict which patients should be referred earlier on, to avoid having them potentially get worse while going through the clinic system? In theory, we could provide such patients with early intensive support, if we were able to identify who they were with reasonable accuracy.

Let's create an ensemble of machine learners, ranging from a simple logistic regression model through a deep learning neural network, to determine whether we can create a predictive model for these patients. A simulated dataset (not real patient data) is available for download on the book website as the file "10_Basu_sampledata.csv", and contains a sample of 50 variables for each of the 2,000 patients in our dataset (reflecting factors like co-morbid diagnoses, previous psychiatric history, and so on).

We can import the data and see that among our 2,000 patients, just under 450 (22%) had to be referred to a psychiatrist (y variable = 1):

```
library(readr)
setwd("~/Downloads")
alldata = read_csv("10_Basu_sampledata.csv")
View(alldata)
summary(alldata)
```

We can split our data into a training dataset and a test dataset to validate our machine learners:

```
library(caret)
```

```
set.seed(1300)
splitIndex <- createDataPartition(alldata$y, p = .8, list = FALSE,
times = 1) # randomly splitting the data into train and test sets
trainSplit <- alldata[ splitIndex,]
testSplit <- alldata[-splitIndex,]
prop.table(table(trainSplit$y))
library(h2o)
h2o.init()
# split up the data into training and testing subsets
train <- as.h2o(trainSplit)
test <- as.h2o(testSplit)
y <- "y"
x <- setdiff(names(train), y)
train[,y] <- as.factor(train[,y])
test[,y] <- as.factor(test[,y])
```

Now suppose are agnostic about what type of machine learner might help us do our predictions. We can simply train an ensemble of models and test how well they perform in the test dataset:

```
aml <- h2o.automl(x = x, y = y,
                  training_frame = train,
                  validation_frame = test,
                  max_runtime_secs = 300)
lb <- aml@leaderboard
lb
aml@leader
```

The `aml@leaderboard` command gets the list of models in the `automl` ensemble and tells us which models performed best in terms of the C-statistic for discrimination:

```
      > lb
                                             model_id       auc   logloss
mean_per_class_error      rmse       mse
1            GBM_grid_0_AutoML_20181011_105258_model_2 0.6511937 0.5068115
0.3956445 0.4044288 0.1635626
2 StackedEnsemble_BestOfFamily_0_AutoML_20181011_105258 0.6501675 0.4991413
0.3762045 0.4015084 0.1612090
3    StackedEnsemble_AllModels_0_AutoML_20181011_105258 0.6478977 0.4980539
0.3899906 0.4007454 0.1605969
4            GBM_grid_0_AutoML_20181011_105258_model_1 0.6439880 0.5068193
0.3935453 0.4045383 0.1636512
5            GBM_grid_0_AutoML_20181011_105258_model_4 0.6439307 0.4997243
0.4056846 0.4017022 0.1613647
6            GBM_grid_0_AutoML_20181011_105258_model_7 0.6382562 0.5254510
0.3997513 0.4090565 0.1673273
```

Here we see that the GBM modeling approach appears to have been the best performer in terms of discrimination, with a C-statistic ('AUC') of 0.65 on the training dataset during internal cross-validation. Next, the `aml@leader` command tells us the performance statistics of the best-performing learner on the validation data:

```
MSE:  0.1792674
RMSE:  0.4233998
LogLoss:  0.5405142
Mean Per-Class Error:  0.3790336
AUC:  0.6393097
Gini:  0.2786195

Confusion Matrix (vertical: actual; across: predicted) for F1-optimal
threshold:
         0   1    Error      Rate
0      138 167 0.547541  =167/305
1       20  75 0.210526    =20/95
Totals 158 242 0.467500  =187/400
```

Here, we see that the C-statistic for this learner is 0.64 on the test data, and that it has an error rate in the confusion matrix of 187/400 = 47%.

That's quite a low discrimination and high error rate, and suggests to me that even though the idea of conducting early referral to a psychiatrist might be a good one, we may not have the right data for the job—some unobserved variables might be influencing who needs a psychiatrist, and we probably need to go back to the clinic and conduct some qualitative interviews or new surveys among practitioners and patients to identify what factors might be relevant to include in our data. The poor performance of even powerful machine learning approaches here provides an important lesson: that simply having a complex model trained on a large dataset with powerful computing technologies is not sufficient to produce a good model. A good model needs to have high-quality data that is relevant to the problem, and our checks on the model in this example reveal that we probably don't have the right data for this problem.

In the next Chapter, we discuss how to determine whether or not a model is truly good or bad, and how even simple models can provide us with useful insights. We discuss how to be a good 'consumer' of models, particularly when complex models are provided to us by consultants or agencies, and we are asked to evaluate such models to determine their believability or applicability to our healthcare or public health system.